A biophysical model of viral escape from polyclonal antibodies

Timothy C. Yu,1,2,3,† Zorian T. Thornton,2,4,† William W. Hannon,1,2,3 William S. DeWitt,2,4 Caelan E. Radford,1,2,3 Frederick A. Matsen IV,3,4,5 and Jesse D. Bloom1,2,4,5,†

1Basic Sciences Division, Fred Hutchinson Cancer Center, 1100 Fairview Ave N, Seattle, WA 98109, USA, 2Computational Biology Program, Fred Hutchinson Cancer Center, 1100 Fairview Ave N, Seattle, WA 98109, USA, 3Molecular and Cellular Biology Graduate Program, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195, USA, 4Department of Genome Sciences, University of Washington, 9720 15th Ave NE, Seattle, WA 98195, USA and 5Howard Hughes Medical Institute, 1100 Fairview Ave N, Seattle, WA 98109, USA

†These authors contributed equally to this work.
†https://orcid.org/0000-0003-1267-3408
*Corresponding author. E-mail: jbloom@fredhutch.org

Abstract

A challenge in studying viral immune escape is determining how mutations combine to escape polyclonal antibodies, which can potentially target multiple distinct viral epitopes. Here we introduce a biophysical model of this process that partitions the total polyclonal antibody activity by epitope and then quantifies how each viral mutation affects the antibody activity against each epitope. We develop software that can use deep mutational scanning data to infer these properties for polyclonal antibody mixtures. We validate this software using a computationally simulated deep mutational scanning experiment and demonstrate that it enables the prediction of escape by arbitrary combinations of mutations. The software described in this paper is available at https://jbloomlab.github.io/polyclonal.

Keywords: viral escape; antibody epitope; deep mutational scanning; biophysical model; epistasis.

Introduction

Many viruses evolve antigenically to escape polyclonal antibodies elicited by vaccination or prior infection (Smith et al. 2004; Hensley et al. 2009; Bedford et al. 2014; Eguia et al. 2021). Neutralization assays are the gold standard for experimentally assessing if a new viral variant has mutations that erode antibody immunity. However, neutralization assays require generating the actual viral variant(s) after they emerge for individual testing, and can therefore only be effectively applied retrospectively to a modest number of viral variants of interest (DeGrace et al. 2022). For this reason, neutralization assays are having difficulty keeping pace with the identification of vast numbers of emerging viral variants by genomic epidemiology (Elbe and Buckland-Merrett 2017; Rambaut et al. 2020; Viana et al. 2022), and so it would be useful to have a method for accurately predicting the antigenic phenotype of viral variants with arbitrary combinations of mutations.

Unfortunately, predicting how a polyclonal serum will neutralize a new viral variant remains a challenge. Deep mutational scanning can systematically measure how large libraries of viral protein variants affect antibody neutralization or binding (Dingens et al. 2017; Lee et al. 2019; Wu et al. 2020; Greaney et al. 2021a, 2021c). However, although such high-throughput experimental methods can assess the effects of all single mutants on some viral proteins, the number of possible multiply mutated variants far exceeds the limits of these experiments. Therefore, a variety of computational approaches have been developed that attempt to predict escape by new viral variants. These approaches include basic transformations of deep mutational scanning data (Greaney, Starr, and Bloom 2022), models that integrate antigenic data with phylogenetic (Neher et al. 2016) or sequence data (Sun et al. 2013; Harvey et al. 2016), and neural networks that can be trained using deep mutational scanning data (Taft et al. 2022) or sequence data alone (Hie et al. 2021; Thadani et al. 2022).

Here we introduce a new model of viral polyclonal antibody escape that has several advantages over existing computational approaches. Our model is interpretable in terms of underlying biophysical parameters, can be directly fit to experimental deep mutational scanning data, and can predict how new viral variants will be neutralized by the polyclonal sera used to generate the experimental data. We implement our model in a software package and validate it on simulated experimental data. Finally, we demonstrate how the parameters of the model provide quantitative intuition for how mutations combine to escape antibodies that target distinct viral epitopes.

Results

The concept of antibody epitopes

Our approach is inspired by the idea that viral antigens can be partitioned into distinct epitopes. This idea can be traced back over four decades to classic experiments on influenza, which tested viral mutants against large panels of monoclonal antibodies (Laver et al. 1979; Yewdell, Webster, and Gerhard 1979;
Viral escape mutants were first selected using individual antibodies and then tested against other antibodies in the panel. A pattern that emerged from these experiments was that groups of antibodies were escaped by similar viral mutants (Fig. 1). Antibodies that share common escape mutants were inferred to recognize a common epitope on the viral protein. For instance, in Fig. 1, Antibodies 2–5 recognize a similar epitope since they are escaped by many of the same viral mutants.

Based on these studies, the H3 influenza hemagglutinin was divided into five distinct epitopes (Wiley, Wilson, and Skehel 1981; Skehel et al. 1984). This type of coarse epitope map also provided a straightforward way to conceptualize viral escape from polyclonal serum. Unlike monoclonal antibodies, polyclonal serum can contain multiple antibodies, each recognizing discrete epitopes. As such, any variant that fully escapes polyclonal serum would need to escape antibodies binding at multiple epitopes.

The concept of dividing viral antigens into distinct epitopes has proven to be very useful and continues to be applied to new viruses such as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Barnes et al. 2020; Piccoli et al. 2020; Cao et al. 2022). However, grouping antibodies by epitopes is a simplifying approximation, as two antibodies targeting a similar epitope may be escaped by slightly different mutations, and some antibodies may bind idiosyncratic regions outside or between major epitopes (Doud, Hensley, and Bloom 2017; Piccoli et al. 2020; Starr et al. 2021a; Greaney et al. 2021a, 2021c). For instance, in Fig. 1, Antibodies 6 and 12 are grouped into the same epitope, even though some viral mutants only escape one of these two antibodies. Nonetheless, the concept of epitopes provides a valuable way to interpret polyclonal antibody escape and forms the basis of the quantitative approach we take here.

**An epitope-based model of viral escape from multiple antibodies**

Given the concept of epitopes described earlier, consider viral escape from multiple antibodies. For simplicity, consider a hypothetical polyclonal mixture of just two antibodies (1 and 2) that bind distinct epitopes on a viral antigen. The functional activities of these two antibodies in the polyclonal mixture can differ due to antibody-intrinsic factors (e.g. binding affinity and neutralization potency) and extrinsic factors (e.g. antibody concentration). In our hypothetical antibody mixture, we assume Antibody 1 has a higher functional activity than Antibody 2.

To understand how mutations affect viral escape from this hypothetical two-antibody mixture, consider the viral variants in Fig. 2. In the first variant, a single mutation escapes Antibody 1 while leaving Antibody 2 binding intact. Since Antibody 1 has a higher functional activity, losing its contribution causes a marked escape from the overall antibody mixture (Fig. 2A). However, a viral variant with a single mutation that escapes Antibody 2 but leaves Antibody 1 binding intact has little escape from the overall antibody mixture, since the more active Antibody 1 can still bind (Fig. 2B).

Importantly, this framework makes distinct predictions about the effects of multiple mutations depending on whether they are in the same or different epitopes. Imagine if the viral variant has two mutations in Antibody 1’s epitope. If one of the mutations is already sufficient to mostly escape binding by Antibody 1, then the second mutation will have a largely redundant effect (Fig. 2C). It follows that multiple escape mutations in the same epitope will have diminishing returns—even when all Antibody 1 molecules are unbound, Antibody 2 can still bind. However, the situation is very different if the viral variant gains mutations in the epitopes of both Antibodies 1 and 2. Since both antibodies are escaped, the polyclonal mixture will have no activity (Fig. 2D). Note that the principles described previously can be readily extended to mixtures of antibodies that target more than two distinct epitopes.

The simple hypothetical example described previously is mirrored in real-world data by Kuzmina et al. (2021) on SARS-CoV-2 escape from neutralization by polyclonal serum pooled from vaccinated individuals (Fig. 3). The K417N mutation, which is located in the subdominant, Class 1 epitope (Greaney et al. 2021b) of the SARS-CoV-2 receptor-binding domain (RBD), has little effect on polyclonal antibody escape. But E484K, which is located in the immunodominant Class 2 epitope (Greaney et al. 2021b), causes a substantial drop in polyclonal antibody neutralization. But while
Figure 2. Viral escape from a hypothetical mixture of antibodies targeting two different epitopes. (A) A single mutation in the dominant epitope substantially decreases but does not eliminate binding by the antibody mixture. (B) A single mutation in the subdominant epitope has a minimal effect on binding by the overall antibody mixture since the binding of antibodies to the dominant epitope is unaffected. (C) Multiple mutations in the dominant epitope have an effect that is no greater than a single mutation in this epitope since the remaining binding is due to antibodies targeting the subdominant epitope. (D) Mutations in both epitopes completely escape binding by the antibody mixture. For each row, the plots show the fraction of antigens unbound by antibody as a function of antibody concentration. The left plot shows the fraction unbound for individual epitopes, and the right plot shows the overall fraction unbound. The dashed lines indicate binding to the unmutated antigen, while solid lines indicate binding to the mutated antigen.

K417N has no effect on its own, combining K417N with E484K causes a larger drop in neutralization than E484K alone, presumably due to escape from two distinct groups of antibodies in the sera targeting each epitope (Fig. 3).

Biophysical modeling of polyclonal antibody escape

Here we formalize the epitope-based model of polyclonal antibody escape described previously in terms of experimental measurable, for instance, from a neutralization assay. Additionally, suppose the antibodies can bind to one of \( E \) epitopes on the viral antigen. Let \( U_e(v, c) \) be the fraction of variant \( v \) that has an unbound epitope \( e \) when the antibody mixture is at concentration \( c \). Assuming antibodies bind independently without competition, we can express \( p(v, c) \) as follows:

\[
p(v, c) = \prod_{e=1}^{E} U_e(v, c).
\]

Note that relaxing the assumption that antibodies bind independently without competition to each epitope has little effect on \( p(v, c) \) (see Appendix). Note also that this model simplifies escape
Virus Evolution

positive values of \( \phi \), let us group together all antibodies targeting a given epitope and let the binding affinity of antibodies binding at another epitope. Furthermore, we assume that there is no competition among antibodies binding to the same epitope, that antibodies to each epitope are completely neutralizing, and that the antibody binding to any given epitope is negligible. A Hill curve with coefficient of 1, then \( U_e(v,c) \) is given by a Hill equation (Einav and Bloom 2020) as follows:

\[
U_e(v,c) = \frac{1 + c \exp(-\phi_e(v))}{1 + c \exp(-\phi_e(v))} = \frac{1}{1 + c \exp(-\phi_e(v))},
\]

where \( -\phi_e(v) \) represents the functional activity of antibodies to epitope \( e \) against variant \( v \). Note that \( \phi_e(v) \) depends on both antibody-intrinsic factors (i.e., affinity or memory) and the free energy of binding \( \Delta G_e(v) \) and extrinsic factors (i.e., relative fraction \( f_e \) of antibodies in the mix that bind epitope \( e \)). The overall functional activity directed to an epitope is a combination of these intrinsic and extrinsic factors, and can be expressed as 

\[
\phi_e(v) \approx \frac{\Delta G_e(v)}{RT} - \ln(f_e).
\]

Note that \( RT \) is the product of the molar gas constant and the temperature, and \( K_{d,e}(v) = \exp\left(\frac{\Delta G_e(v)}{RT}\right) \) is the dissociation constant. For simplicity, we consider each \( \phi_e(v) \) to be independent, so antibody binding at one epitope does not affect the affinity of antibodies binding at another epitope. Furthermore, we group together all antibodies targeting a given epitope and let \( \phi_e(v) \) be their effective activity, simplifying the fact that each of these antibodies can have different activities and affinities. More positive values of \( -\phi_e(v) \) indicate that antibodies that bind epitope \( e \) have higher functional activity against variant \( v \).

Lastly, we can write \( \phi_e(v) \) in terms of the contributions of specific mutations. To do this, assume that mutations have additive effects on the free energy of binding for antibodies targeting any given epitope \( e \) (Otwinowski, McCandlish, and Plotkin 2018; Otwinowski and Wilke 2018). In this way, we assume that antigenic epistasis arises via the non-linear Hill function that relates \( \phi_e(v) \) to \( U_e(v,c) \) and the product over \( U_e(v,c) \) across epitopes. Specifically, let \( a_{uni,e} \) be the functional activity of antibodies on the unmutated viral antigen, with larger values of \( a_{uni,e} \) indicating stronger antibody activity targeting this epitope. Let \( \beta_{m,e} \) be the extent to which mutation \( m \) reduces antibody binding or neutralization at epitope \( e \), with larger values of \( \beta_{m,e} \) corresponding to a larger contribution to antibody escape. So, \( \phi_e(v) \) can also be written as:

\[
\phi_e(v) = -a_{uni,e} + \sum_{m=1}^{M} \beta_{m,e} b(v)_m,
\]

where \( b(v)_m \) is one if variant \( v \) has mutation \( m \) and 0 otherwise, and \( m \) ranges over all \( M \) possible mutations. Taken together, the above equations relate the pre-mutation functional activity of antibodies at each epitope \( a_{uni,e} \) and the antibody escape effects of individual mutations at each epitope \( \beta_{m,e} \) to the experimentally measured fraction \( p(v,c) \) of variants \( v \) that escape an antibody mixture at concentration \( c \).

**Fitting biophysical models to deep mutational scanning data**

We reasoned that the parameters of our biophysical model could be learned from experiments that measure \( p(v,c) \) for a large number of variants, containing all possible single mutants and a sizable number of multiple mutants (Fowler et al. 2010; Kinney et al. 2010). These rich genotype–phenotype measurements can be obtained by viral deep mutational scanning. Briefly, viral deep mutational scanning involves generating a large library of viral or viral protein variants, incubating this library with antibodies or sera, and using deep sequencing to identify which variants successfully escape binding or neutralization (Dingens et al. 2017; Lee et al. 2019; Wu et al. 2020; Greaney et al. 2021a; Greaney et al. 2021c). While previous studies have been restricted to variants that predominantly contain single amino acid mutations, recent advances enable combinations of mutations to be assayed as well (Dadonnaite et al. 2022). Although only an infinitesimal fraction of all combinations of mutations can be assayed, this approach does measure \( p(v,c) \) for each mutation in many different backgrounds, which should be sufficient for revealing the epitopes targeted by polyclonal antibody mixtures and their associated escape mutations.

To fit our biophysical model using deep mutational scanning data, we created a Python software package named polyclonal that uses gradient-based optimization to fit the model to a large set of viral variants \( v \) and their corresponding experimentally measured escape values, \( p(v,c) \). This software package estimates the \( a_{uni,e} \) and \( \beta_{m,e} \) parameters that best predict the measured \( p(v,c) \) under tunable and biologically motivated constraints. These constraints are key to ensuring the inference of biologically accurate epitopes. We typically enforce a sparsity constraint that encourages most \( \beta_{m,e} \) values to be close to 0, as most mutations to a

**Figure 3.** Mutations at multiple epitopes have a synergistic effect on viral escape in real experimental data. (A) Structure of the SARS-CoV-2 RBD highlighting locations of the K417N and E484K mutations in the Class 1 and 2 epitopes, respectively. (B) Neutralization of the unmutated SARS-CoV-2 spike, the K417N and E484K single mutants, and the double mutant against pooled serum from vaccinated individuals. Note how K417N alone has little effect on neutralization but does cause substantial additional escape in the background of E484K. Experimental data were taken from Kuzmina et al. (2021) and replotted here.
viral antigen should not mediate antibody escape. Similarly, the $a_{\text{wt,e}}$ values are penalized to be close to 0 unless there is clear evidence to the contrary. We also usually enforce an evenness constraint, as most mutations to a site that mediates antibody escape at a specific epitope tend to have similar effects (or $\beta_{m,e}$’s). The software itself is available at https://github.com/jbloomlab/polyclonal, and detailed documentation is at https://jbloomlab.github.io/polyclonal. The software also provides methods to visualize the resulting mutation-level escape values in interactive plots, as described in the documentation.

Validation with computationally simulated deep mutational scanning data

To validate the performance of our software, we simulated a hypothetical polyclonal antibody mixture that contains antibodies targeting three major neutralizing epitopes (Class 1–3) on the SARS-CoV-2 RBD (Fig. 4A) (Barnes et al. 2020; Greaney et al. 2021b). We assigned each epitope a different pre-mutation functional activity ($a_{\text{wt,e}}$) based on the order of typical neutralizing activities against these epitopes in actual human polyclonal serum (Class 2 > Class 3 > Class 1) (Greaney et al. 2021b) (Fig. 4B). Next, we assigned $\beta_{m,e}$ values using prior deep mutational scanning studies (Starr et al. 2021b, 2021c) that measured the effects of all single RBD mutations on escape from prototypical monoclonal antibodies targeting each epitope (Fig. 4C). Specifically, we used LY-CoV016 (etesevimab) for the prototype Class 1 antibody, LY-CoV555 (bamlanivimab) for Class 2, and REGN10987 (imdevimab) for Class 3 (Starr et al. 2021b; Starr et al. 2021c). Note that the site of greatest total escape (the sum of positive $\beta_{m,e}$ values at a site) is 417 for Class 1, 484 for Class 2, and 444 for Class 3.

Based on this hypothetical polyclonal antibody mixture, we computationally simulated a realistic deep mutational scanning dataset containing 30,000 RBD variants. These variants contained an average of two amino acid mutations, with the number of mutations per variant following a Poisson distribution (Supplementary Fig. S1A). These variants also only contained mutations at sites found to be functionally tolerated in prior deep mutational scanning, and these sites were well-represented in the simulated dataset (Supplementary Fig. S1B). We then calculated the true $p(v,c)$ for each variant using our biophysical model with the $a_{\text{wt,e}}$ and $\beta_{m,e}$ defined earlier. We did this for three concentrations of the hypothetical serum that represent IC57.5, IC99.9, and IC99.998 against the unmutated RBD (Supplementary Fig. S1C).

Lastly, we added gaussian noise, $N(0,0.05)$, into the $p(v,c)$ measurements and then truncated the noisy values to be between 0 and 1 to reflect experimental errors and built in a 1 percent probability of adding or subtracting a mutation from a variant to portray sequencing errors.

We found that polyclonal could successfully infer the underlying properties of our hypothetical polyclonal antibody mixture when fit to the noisy, simulated deep mutational scanning dataset. The predicted $a_{\text{wt,e}}$ values were nearly identical to the true $a_{\text{wt,e}}$ values (Fig. 5A), suggesting that we can deconvolve the dominance hierarchy of epitopes targeted by a polyclonal antibody mixture. Furthermore, the predicted $\beta_{m,e}$ values strongly correlated with the true $\beta_{m,e}$ values (Fig. 5B, $R^2 = 0.63$ for Class 1, $R^2 = 0.91$ for Class 2, and $R^2 = 0.85$ for Class 3), indicating that we can learn the effects of mutations on antibody escape at each epitope. Note that the lower correlation for the Class 1 epitope can be attributed to its relative subdominance. As shown in Fig. 2B, mutations to a subdominant epitope manifest little effect on the functional activity of the overall antibody mixture when the immunodominant epitopes remain unaffected, and so only have measurable effects in the subset of mutants with mutations in more dominant epitopes. Nonetheless, our approach can still learn these subdominant effects reasonably well. Taken together, these results demonstrate that noisy $p(v,c)$’s measured by deep mutational scanning can be modeled to reveal fine details of polyclonal antibody mixtures: the extent to which antibodies target specific epitopes on a viral antigen and the extent to which mutations escape antibodies to each epitope.

We used the fit biophysical model to predict the extent to which variants that were not seen in our experiments will escape the same polyclonal antibody mixture. To do this, we simulated an independent deep mutational scanning dataset, grounded in the same true $a_{\text{wt,e}}$ and $\beta_{m,e}$ values, but with a higher number of mutations (three) on average per variant (Supplementary Fig. S1A). We found that our fit model could predict the IC90...
Virus Evolution

Figure 5. Model validation using a computationally simulated deep mutational scanning dataset. (A) Pre-mutation functional activities of antibodies ($a_{wt,e}$) inferred by the fit model match the actual $a_{wt,e}$ for each epitope. (B) Mutation escape effects ($\beta_{m,e}$) inferred by the fit model strongly correlate with the actual $\beta_{m,e}$ for each epitope used in the simulation. Mutation escape effects for the immunodominant Class 2 epitope are most correlated, while escape effects for the subdominant Class 1 epitope are least correlated. (C) The IC90’s predicted by the fit model strongly correlate with the actual IC90’s of an independently simulated dataset with a higher mutation rate (three mutations on average per variant).

Model fitting is dependent on experimental design

We next sought to clarify the experimental conditions for deep mutational scanning that lead to accurate model fitting with polyclonal. To that end, we systematically explored the impact of three important experimental conditions on model fitting through simulation (see Methods).

We first examined the effect of the library mutation rate, an important consideration for library design because variants containing multiple mutations are key to revealing epitopes. If a library only contained variants with single mutations, it should not be possible to detect the presence of subdominant epitopes, due to the phenomenon in Fig. 2B. Indeed, we found that a library containing 30,000 variants with an average of one mutation could only infer the $\beta_{m,e}$ values for the immunodominant Class 2 epitope (Fig. 6A), highlighting the need for including multiply mutated variants. On the other hand, a library containing 30,000 variants with two mutations on average was sufficient for inferring the $\beta_{m,e}$ values at all three epitopes, but the accuracy improved as the mutation rate increased (Fig. 6A), consistent with our expectation that greater coverage of variants with multiple epitope mutations is helpful. Note that these simulations do not capture the real-world fact that variants will be increasingly less likely to be functional as they accumulate more mutations.

Next, we investigated the effect of library size, another important library design consideration. If there are too few variants in the library, mutations may not be observed in enough backgrounds to be able to deconvolve their effects on different epitopes. Previously, we showed that a simulated library with 30,000 variants with three mutations on average could accurately infer the $\beta_{m,e}$ values at all three epitopes. Therefore, we set out to determine the minimum number of variants, with three mutations on average, required to accurately infer the $\beta_{m,e}$ values at all three epitopes. We found that at least 20,000 functional variants were required to accurately infer the $\beta_{m,e}$ values at each epitope (Fig. 6B). We also noticed that the $\beta_{m,e}$ values of immunodominant epitopes can be inferred with fewer variants (Fig. 6B), consistent with the idea that effects of mutations at subdominant epitopes can only be observed in the backgrounds where the immunodominant epitope is already mutated.

Lastly, we explored the effect of antibody concentration, an important consideration for deep mutational scanning selections. If the concentration is too high, all variants will be neutralized. If the concentration is too low, there will be little measurable effect from the mutations. To test the effect of different concentrations, we again used the simulated library containing 30,000 variants with three mutations on average. First, we tested if the data collected at a single concentration was sufficient. We found that a single concentration of IC99.9 against the unmutated RBD was most effective at inferring the $\beta_{m,e}$ values for each epitope, but there does indeed exist a fine balance (Fig. 6C). Particularly for the
subdominant Class 1 epitope, the accuracy is lower when the concentration is too low or too high. We then tested if model fitting could be improved by including additional concentrations flanking the IC99.9, spanning the IC91.4 to the IC99.998. Indeed, we found that the accuracy increases as the number of concentrations measured increases (Fig. 6D), consistent with our expectation that obtaining more data points along the binding curve is helpful for quantifying the effects of mutations.

Discussion

We have described a biophysical model of viral escape from polyclonal antibodies. Notably, this model is composed of easily interpretable parameters that capture the interactions between antibodies and the viral epitopes they bind. In addition, we developed a software package that can infer these parameters using deep mutational scanning data. Using a simulated deep mutational scanning dataset, we demonstrated that this approach can infer true parameters from noisy experiments if the deep mutational scanning experiment is appropriately designed.

There are several limitations to our approach. Grouping antibodies by discrete epitopes is an approximation because each antibody is unique; however, decades of experimental work have shown that an epitope-based representation offers a useful way to interpret viral escape. Our model also does not explicitly consider the fact that there are multiple antigens per virion and instead models escape as simply being the state where no antibodies bind a single idealized antigen. Additionally, our model assumes a specific shape of antigenic epistasis: mutations have additive effects on binding affinity at each epitope but can manifest non-linear effects on the overall measured phenotype (e.g., neutralization) due to both the non-linear Hill curves that relate binding affinity to the total fraction bound at an epitope and the product of these fractions across epitopes. Note that the single-epitope version of our model is similar to the global epistasis models that have proven so useful for interpreting some other types of deep mutational scanning data (Otwinowski, McCandlish, and Plotkin 2018; Tareen et al. 2022), while the multi-epitope version resembles a fuzzy logic model of multiple biophysical traits (Warszawski et al. 2014). Our model also assumes that antibody binding to one epitope does not influence the affinity of antibodies to other epitopes and that the Hill curves have a coefficient of one. It is possible that these assumptions could be relaxed in elaborated versions of our model. We expect that these assumptions will hold reasonably well for viral variants with modest numbers of mutations but may not extrapolate accurately to variants with dozens of mutations relative to the parental strain used for the deep mutational scanning experiment.

Despite these limitations, our model not only predicts the escape potential of viral variants with arbitrary combinations of mutations that are observed in a simulated deep mutational scanning library—it also clarifies how escape mutations combine to determine the magnitude of viral escape. While this manuscript was under review, our model was successfully applied to real deep mutational scanning data measuring the escape of SARS-CoV-2 spike variants from monoclonal antibodies, yielding predictions...
that strongly correlated with neutralization assays (Dadonaite et al. 2022). While this paper only considers simulated data, we envision that our model can be further applied to appropriately designed deep mutational scanning experiments to address two main questions: (1) delineating the epitopes targeted by polyclonal serum and (2) predicting the antigenic properties of new variants with arbitrary combinations of mutations.

Methods

Model fitting

Our goal is to estimate the biophysical model parameters that best predict the deep mutational scanning escape measurements under biologically motivated constraints. We defined a loss function that is robust to outliers:

\[ L = \sum_{v,c} h_{\text{loss}}(p(v,c) - \hat{p}(v,c)), \]

where \( p(v,c) \) is the actual antibody escape fraction for variant \( v \) at concentration \( c \) and \( \hat{p}(v,c) \) is the predicted antibody escape fraction for variant \( v \) at concentration \( c \), and \( h_{\text{loss}}(x) \) is a scaled Pseudo-Huber function, defined as follows:

\[ h_{\text{loss}}(x) = \frac{\delta^2}{\delta} \left( \frac{1 + (x/\delta)^2}{2} - 1 \right), \]

where \( \delta \) is a parameter that indicates when the loss transitions from being quadratic (L2-like) to linear (L1-like). Note that the loss is mostly L2-like when \( x \), the residual, is small. However, the loss transitions to become more L1-like when \( x \) is large, making it more robust to outliers. A default \( \delta_{\text{loss}} = 0.01 \) was used in all fit models. To minimize this loss function, we use the gradient-based L-BFGS-B method implemented in scipy.optimize.

Regularization

Additionally, we implemented penalty terms to regularize the parameters to behave under biologically motivated constraints. We regularize the escape values \( \beta_{m,e} \) using a Pseudo-Huber function, based on the notion that most mutations should not mediate escape.

\[ R_{\text{escape}} = \lambda_{\text{escape}} \sum_{m,e} h_{\text{escape}}(\beta_{m,e}), \]

where \( \lambda_{\text{escape}} \) is the strength of the regularization and \( \delta_{\text{escape}} \) is the Pseudo-Huber delta parameter. For all fit models, the default parameters \( \lambda_{\text{escape}} = 0.02 \) and \( \delta_{\text{escape}} = 0.1 \) were used.

We also regularize the variance of escape values \( \beta_{m,e} \) at each site, based on the notion that mutations at a site involved in escape will exhibit similar effects.

\[ R_{\text{spread}} = \lambda_{\text{spread}} \sum_{e,i} \left( \beta_{m,e} - \frac{1}{M_i} \sum_{m \in \xi} \beta_{m',e} \right)^2, \]

where \( \lambda_{\text{spread}} \) is the strength of the regularization, \( i \) ranges over all sites, \( M_i \) is the number of mutations at site \( i \), and \( m \in \xi \) indicates all mutations at site \( i \). For all models fit, the default parameter \( \lambda_{\text{spread}} = 0.25 \) was used.

Lastly, we regularize the pre-mutation functional activities \( a_{w,t,e} \) using a Pseudo-Huber function as these should be less positive (or even negative) unless there is clear evidence to the contrary.

\[ R_{\text{activity}} = \lambda_{\text{activity}} \sum_{e} h_{\text{activity}}(a_{w,t,e}) \]

if \( a_{w,t,e} \geq 0 \)

where \( \lambda_{\text{activity}} \) is the strength of the regularization, \( \delta_{\text{activity}} \) is the Pseudo-Huber delta parameter, and \( R_{\text{activity}} = 0 \) when \( a_{w,t,e} < 0 \). For all models fit, the default parameters \( \lambda_{\text{activity}} = 1.0 \) and \( \delta_{\text{activity}} = 0.1 \) were used.

For more information on these penalties and model fitting, see https://jbloomlab.github.io/polyclonal/optimization.html.

Determining the number of epitopes

Fitting the model requires specifying the number of epitopes a priori. However, it is not possible to know the number of epitopes that are targeted by antibodies in polyclonal serum in practice. Our approach to resolving this is similar to the `elbow method` commonly used to determine the optimal number of clusters in k-means clustering. We start by fitting a model with one epitope and iteratively fit models with an increasing number of epitopes. At some point, the \( N \)-th epitope becomes redundant. This is evidenced by a highly negative \( a_{w,t,e} \) value (i.e. if antibodies existed against this epitope, they would never be bound) and all near-zero \( \beta_{m,e} \) values, indicating that the previous fit model, containing \( N - 1 \) epitopes, is the one that best describes the polyclonal mixture. To view how this approach was applied to the simulated RBD example, see https://jbloomlab.github.io/polyclonal/simulate_RBD.html.

Experimental design simulation

We computationally simulated deep mutational scanning libraries based on the hypothetical polyclonal antibody mixture shown in Fig. 4. All libraries contained 30,000 variants but differed by their mutation rate with variants containing an average of one, two, three, or four mutations. Furthermore, the number of mutations per variant in each library followed a Poisson distribution. Variant escape was also simulated under six concentrations in each library. These concentrations represented IC91.443, IC97.488, IC99.441, IC99.95, IC99.985, and IC99.98 against the unmutated RBD antigen. For more details on how these experiments were simulated, see https://jbloomlab.github.io/polyclonal/simulate_RBD.html.

To make comparisons about model fitting, models were fit with identical parameters except for the experimental variable of interest: library mutation rate, library size, and antibody concentration.

To determine the impact of library mutation rate, models were fit to simulated datasets for libraries containing one, two, three, or four mutations on average. All datasets contained 30,000 variants and were measured at three different concentrations.

To determine the impact of library size, models were fit to simulated datasets containing different-sized subsets of variants that were randomly sampled from a library containing an average of three mutations per variant and measured at three different concentrations.

To determine the impact of antibody concentration, models were fit to simulated datasets to a library measured at one or multiple antibody concentrations. This library contained 30,000 variants and three mutations on average per variant.

For more information, see https://jbloomlab.github.io/polyclonal/expt_design.html.
Code and data availability

Software source code, along with code and data that reproduce the figures (except Fig. 3), are available at: https://github.com/jbloomlab/polyclonal
Code and data for reproducing the neutralization curves in Fig. 3 are available at: https://github.com/jbloomlab/polyclonal-paper
Software documentation is available at: https://jbloomlab.github.io/polyclonal

Supplementary data

Supplementary data are available at Virus Evolution online.

Acknowledgements

This work was supported in part by the National Institutes of Health/National Institute of Allergy and Infectious Diseases under contract 75N9321C00015 (to J.D.B.), grants R01AI141707 and R01AI165821 (to J.D.B.), and grant FI3AI150163 (to W.S.D.). The work was also supported by the National Institutes of Health/National Institutes of General Medical Sciences (Cell and Molecular Biology Training Grant (T32 GM007270) to T.C.Y. and the National Science Foundation Graduate Research Fellowship (DGE-2140004) to T.C.Y. J.D.B. and F.A.M. are investigators of the Howard Hughes Medical Institute.

Conflict of interest: J.D.B. consults for Apriori Bio and Oncorus on topics related to viruses or vaccines and recently consulted for Moderna and Merck on similar topics. J.D.B. and C.E.R. receive a share of the intellectual property as inventors on Fred Hutch licensed patents related to viral deep mutational scanning.

References

Barnes, C. O. et al. (2020) ‘SARS-CoV-2 Neutralizing Antibody Structures Inform Therapeutic Strategies’, Nature, 588: 682–7.
Bedford, T. et al. (2014) ‘Integrating Influenza Antigenic Dynamics with Molecular Evolution’, eLife, 3: e01914.
Cao, Y. et al. (2022) ‘Omicron Escapes the Majority of Existing SARS-CoV-2 Neutralizing Antibodies’, Nature, 602: 657–63.
Dadonaite, B. et al. (2022) ‘A Pseudovirus System Enables Deep Mutational Scanning of the Full SARS-CoV-2 Spike’, bioRxiv, 202210.13.512056.
DeGrace, M. M. et al. (2022) ‘Defining the Risk of SARS-CoV-2 Variants on Immune Protection’, Nature, 605: 640–52.
Dingens, A. S. et al. (2017) ‘Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody’, Cell Host & Microbe, 21: 777–87.e4.
Doud, M. B., Hensley, S. E., and Bloom, J. D. (2017) ‘Complete Mapping of Viral Escape from Neutralizing Antibodies’, PLoS Pathogens, 13: e1006271.
Eguia, R. T. et al. (2021) ‘A Human Coronavirus Evolves Antigenically to Escape Antibody Immunity’, PLoS Pathogens, 17: e1009453.
Einav, T. and Bloom, J. D. (2020) ‘When Two are Better than One: Modeling the Mechanisms of Antibody Mixtures’, PLoS Computational Biology, 16: e1007830.
Elbe, S., and Buckland-Merrett, G. (2017) ‘Data, Disease and Diplomacy: GISAID’s Innovative Contribution to Global Health: Data, Disease and Diplomacy’, Global Challenges, 1: 33–46.
Fowler, D. M. et al. (2010) ‘High-resolution Mapping of Protein Sequence-function Relationships’, Nature Methods, 7: 741–6.
Greaney, A. J. et al. (2021a) ‘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.
Greaney, A. J. et al. (2021b) ‘Mapping Mutations to the SARS-CoV-2 RBD that Escape Binding by Different Classes of Antibodies’, Nature Communications, 12: 4196.
Greaney, A. J. et al. (2021c) ‘Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain That Escape Antibody Recognition’, Cell Host & Microbe, 29: 44–57.e9.
Greaney, A. J., Starr, T. N., and Bloom, J. D. (2022) ‘An Antibody-escape Estimator for Mutations to the SARS-CoV-2 Receptor-binding Domain’, Virus Evolution, 8: vea021.
Harvey, W. T. et al. (2016) ‘Identification of Low- and High-impact Hemagglutinin Amino Acid Substitutions That Drive Antigenic Drift of Influenza A(H1N1) Viruses’, PLoS pathogens, 12: e1005526.
Hensley, S. E. et al. (2009) ‘Hemagglutinin Receptor Binding Aviridity Drives Influenza A Virus Antigenic Drift’, Science, 326: 734–6.
Hie, B. et al. (2021) ‘Learning the Language of Viral Evolution and Escape’, Science, 371: 284–8.
Kinfen, J. B. et al. (2010) ‘Using Deep Sequencing to Characterize the Biophysical Mechanism of a Transcriptional Regulatory Sequence’, Proceedings of the National Academy of Sciences of the United States of America, 107: 9158–63.
Kuzmina, A. et al. (2021) ‘SARS-CoV-2 Spike Variants Exhibit Differential Infectivity and Neutralization Resistance to Convalescent or Post-vaccination Sera’, Cell Host & Microbe, 29: 522–528.e2.
Laver, W. G. et al. (1979) ‘Antigenic Drift in Type A Influenza Virus: Sequence Differences in the Hemagglutinin of Hong Kong (H3N2) Variants Selected with Monoclonal Hybridoma Antibodies’, Virology, 98: 226–37.
Lee, J. M. et al. (2019) ‘Mapping Person-to-person Variation in Viral Mutations That Escape Polyclonal Serum Targeting Influenza Hemagglutinin’, eLife, 8: e49324.
Neher, R. A. et al. (2016) ‘Prediction, Dynamics, and Visualization of Antigenic Phenotypes of Seasonal Influenza Viruses’, Proceedings of the National Academy of Sciences of the United States of America, 113: E1701–9.
Otwinowski, J., McCandlish, D. M., and Plotkin, J. B. (2018) ‘Inferring the Shape of Global Epistasis’, Proceedings of the National Academy of Sciences of the United States of America, 115: E7550–8.
Otwinowski, J., and Wölke, C. (2018) ‘Biophysical Inference of Epistasis and the Effects of Mutations on Protein Stability and Function’, Molecular Biology and Evolution, 35: 2345–54.
Piccoli, L. et al. (2020) ‘Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-guided High-Resolution Serology’, Cell, 183: 1024–42.e21.
Rambaut, A. et al. (2020) ‘A Dynamic Nomenclature Proposal for SARS-CoV-2 Lineages to Assist Genomic Epidemiology’, Nature Microbiology, 5: 1403–7.
Skehel, J. J. et al. (1984) ‘A Carbohydrate Side Chain on Hemagglutinins of Hong Kong Influenza Viruses Inhibits Recognition by a Monoclonal Antibody’, Proceedings of the National Academy of Sciences of the United States of America, 81: 1779–83.
Smith, D. J. et al. (2004) ‘Mapping the Antigenic and Genetic Evolution of Influenza Virus’, Science, 305: 371–6.
Starr, T. N. et al. (2021a) ‘SARS-CoV-2 RBD Antibodies That Maximize Breadth and Resistance to Escape’, Nature, 597: 97–102.
Starr, T. N. et al. (2021b) ‘Prospective Mapping of Viral Mutations That Escape Antibodies Used to Treat COVID-19’, Science, 371: 850–4.
Starr, T. N. et al. (2021c) ‘Complete Map of SARS-CoV-2 RBD Mutations That Escape the Monoclonal Antibody LY-CoV555 and Its Cocktail with LY-CoV016’, Cell Reports Medicine, 2: 100255.
Sun, H. et al. (2013) 'Using Sequence Data to Infer the Antigenicity of Influenza Virus', mBio, 4: e00230–13.

Tauf, J. M. et al. (2022) 'Deep Mutational Learning Predicts ACE2 Binding and Antibody Escape to Combinatorial Mutations in the SARS-CoV-2 Receptor Binding Domain', Cell 185: 4008–22.

Tareen, A. et al. (2022) 'MAVE-NN: Learning Genotype-phenotype Maps from Multiplex Assays of Variant Effect', Genome Biology, 23: 98.

Thadani, N. N. et al. (2022) 'Learning from Pre-Pandemic Data to Forecast Viral Antibody Escape', bioRxiv. 2022.07.21.501023.

Viana, R. et al. (2022) 'Rapid Epidemic Expansion of the SARS-CoV-2 Omicron Variant in Southern Africa', Nature, 603: 679–86.

Warszawski, S. et al. (2014) 'A "Fuzzy"-logic Language for Encoding Multiple Physical Traits in Biomolecules', Journal of Molecular Biology, 426: 4125–38.

Webster, R. G., and Laver, W. G. (1980) 'Determination of the Number of Nonoverlapping Antigenic Areas on Hong Kong (H3N2) Influenza Virus Hemagglutinin with Monoclonal Antibodies and the Selection of Variants with Potential Epidemiological Significance', Virology, 104: 139–48.

Wiley, D. C., Wilson, I. A., and Skehel, J. J. (1981) 'Structural Identification of the Antibody-binding Sites of Hong Kong Influenza Haemagglutinin and Their Involvement in Antigenic Variation', Nature, 289: 373–8.

Wu, N. C. et al. (2020) 'Different Genetic Barriers for Resistance to HA Stem Antibodies in Influenza H3 and H1 Viruses', Science, 368: 1335–40.

Yewdell, J. W., Webster, R. G., and Gerhard, W. U. (1979) 'Antigenic Variation in Three Distinct Determinants of an Influenza Type A Haemagglutinin Molecule', Nature, 279: 246–8.

Appendix

We have described a biophysical model that assumes a polyclonal antibody mixture can be divided into independent groups of antibodies that bind to distinct epitopes without competition. Here we introduce the validity of this assumption, being cognizant of the observation that realistic viral epitopes are often overlapping and therefore not distinct. To do this, we draw from statistical mechanics principles to compare the antibody escape fractions predicted by our independent epitope model and an identically formulated model that instead assumes all epitopes are overlapping.

1. Monoclonal antibody case

Before considering the polyclonal antibody case, we first consider the case of a monoclonal antibody that binds a viral antigen. Here, the viral protein can exist in two microstates: bound or unbound by the antibody. The Boltzmann weight is 1 for the unbound state and $\frac{1}{K_d}$ for the bound state, where $c_i$ is the antibody concentration and $K_d$ is the dissociation constant of antibodies binding to epitope $e$.

These weights can be derived using the steady-state approximation (Eina, Bloom, and Antia 2020). We can then define the partition function $\Xi$ as follows:

$$\Xi = \sum_i Z_i = Z_{\text{unbound}} + Z_{\text{bound}}$$

where $Z_i$ represents the Boltzmann weights of the $i$ microstates. Given this, the probability of a viral antigen being unbound is

$$p_{\text{unbound}} = \frac{Z_{\text{unbound}}}{\Xi} = \frac{1}{1 + \frac{c_i}{K_d}}$$

2. Polyclonal antibody case

For a polyclonal antibody mixture, we modify $e$ to represent the concentration of the polyclonal antibody mixture. We assume that the polyclonal antibody mixture contains antibodies that bind one of $E$ epitopes. As follows, the Boltzmann weight of the state where epitope $e$ is bound is modified to $\frac{f_e}{K_{d,e}}$, where $f_e$ represents the fraction of antibodies in the mixture that target epitope $e$, and $K_{d,e}$ is the dissociation constant of antibodies binding to epitope $e$.

2.1 Two distinct epitopes

In a polyclonal antibody mixture, new microstates exist where multiple epitopes are bound by antibodies. For example, we can consider a viral antigen that contains two distinct epitopes (1 and 2) that are targeted by polyclonal antibodies. In addition to the microstates where a single epitope is bound, we now require an additional microstate where both epitopes are bound. The Boltzmann weight for this new microstate is:

$$Z_{12,\text{bound}} = \left( \frac{f_1}{K_{d,1}} \right) \left( \frac{f_2}{K_{d,2}} \right)$$

We can then rewrite the partition function $\Xi$ as follows:

$$\Xi = \sum_i Z_i = Z_{\text{unbound}} + Z_{1,\text{bound}} + Z_{2,\text{bound}} + Z_{12,\text{bound}}$$

and the probability of a viral antigen being unbound is

$$p_{\text{unbound}} = \frac{Z_{\text{unbound}}}{\Xi} = \frac{1}{\left( 1 + \frac{c_1}{K_{d,1}} + \frac{c_2}{K_{d,2}} + \left( \frac{f_1}{K_{d,1}} \right) \left( \frac{f_2}{K_{d,2}} \right) \right)}$$

Note that this is the biophysical model that is described in the main text.

2.2 Two overlapping epitopes

In Appendix section 2.1, the two epitopes were distinct, and there was no competition among antibodies. However, if the epitopes are overlapping and there is competition, then the microstate where both epitopes are bound ($Z_{12,\text{bound}}$) can no longer exist. In this case, the probability of a viral antigen being unbound is

$$p_{\text{unbound}} = \frac{Z_{\text{unbound}}}{\Xi} = \frac{1}{\left( 1 + \frac{c_1}{K_{d,1}} + \frac{c_2}{K_{d,2}} \right)}$$

2.3 Extending beyond two epitopes

The same logic applies to viral antigens with more than two epitopes targeted by antibodies. For example, we can write $p_{\text{unbound}}$ for the case of three distinct epitopes as:

$$p_{\text{unbound}} = \frac{Z_{\text{unbound}}}{\Xi} = \frac{1}{\left( 1 + \frac{c_1}{K_{d,1}} + \frac{c_2}{K_{d,2}} + \left( \frac{f_1}{K_{d,1}} \right) \left( \frac{f_2}{K_{d,2}} \right) \right)}$$

and the case of three overlapping epitopes as:

$$p_{\text{unbound}} = \frac{Z_{\text{unbound}}}{\Xi} = \frac{1}{\left( 1 + \frac{c_1}{K_{d,1}} + \frac{c_2}{K_{d,2}} + \frac{c_3}{K_{d,3}} \right)}$$

Overall, we found that the predicted $p_{\text{unbound}}$ does not differ by much between the independent and overlapping epitope models under realistic $c_i$’s, $f_i$’s, and $K_{d,i}$’s. However, the predicted $p_{\text{unbound}}$
becomes more discordant between the two models as the number of epitopes increases, given the number of microstates with multiply bound epitopes increases. In practice, we only fit models to a handful of epitopes and not all of them will overlap. As such, we maintain that the independent epitope assumption is appropriate. To interactively explore how $p_{\text{unbound}}$ differs between the two models, see https://jbloomlab.github.io/polyclonal/partition_function.html.