A lentiviral vector deep mutational scanning system for studying virus evolution and escape from neutralization by antibodies and polyclonal serum

Caelan Radford

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Jesse Bloom, Chair

Julie Overbaugh

Neil King

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

A lentiviral vector deep mutational scanning system for studying virus evolution and escape from neutralization by antibodies and polyclonal serum

Caelan Radford

Chair of the Supervisory Committee:

Jesse D. Bloom

Fred Hutchinson Cancer Research Center

Viral entry proteins are critical for viral replication of enveloped viruses. These proteins allow viruses to bind receptors on cells and accomplish membrane fusion. Since viral entry proteins are located on the viral surface and the surface of infected cells, they are also the targets of host immune responses. For all of these reasons, mutations to viral entry protein can have a wide range of effects. These effects include the ability to better infect new host species, changes to receptor tropism and affinity, and evasion of host immune responses. By studying the effects of mutations to viral entry proteins, we can inform monitoring of emerging viral pathogens as well as the design of vaccines and other therapeutics.

Studying the effects of mutations on viral entry proteins is not straightforward. Traditional methods of measuring mutation effects on viral entry proteins have drawbacks. Often, mutations are cloned individually or in a few combinations and tested in viral replication or neutralization assays, which can be time and resource expensive. These methods may rely on using replicative viruses with full-length genomes, raising the biosafety level required to perform these

experiments. Mutations also often have different effects in different strains of a virus, making it difficult to gain a full picture of the mutation's potential effects unless these already intensive experiments are extended to multiple strains. Some of these challenges have been alleviated by high throughput techniques like phage display, yeast display, deep mutational scanning of full-length replicative viruses, and cell surface display. However, these methods each have drawbacks of their own, such as not displaying the full viral entry proteins, only being able to measure ligand or antibody binding rather than viral entry protein function, not being able to measure effects of combinations of mutations, or being difficult to perform using certain viruses due to technical reasons or biosafety concerns.

We have developed a lentiviral vector-based system for deep mutational scanning of viral entry proteins that overcomes many limitations of previous studies. Non-replicative lentiviral vectors can be pseudotyped by displaying viral entry proteins from viruses of interest on their surface. We developed a method to generate a genotype-phenotype link between a lentivirus genome and a viral entry protein mutant displayed on the surfaces of the virion for large mutant libraries of viral entry proteins. These mutant lentivirus libraries can then be used to measure mutation effects on function and immune escape of full viral entry proteins. By using a nucleotide barcoding method, this system is also able to measure the effects of combinations of mutations in viral entry proteins. Measurements of effects of combinations of mutations allows us to investigate epistasis in mutation effects and map virus escape from immunity targeting multiple epitopes simultaneously.

We used the lentiviral vector system to perform deep mutational scanning of the SARS-CoV-2 spike protein. This allowed us to safely measure the effects of mutations to the full SARS-CoV-2 Spike protein in a biosafety level 2 setting. Using mutant libraries of Omicron BA.1 and Delta strain spike proteins, we mapped escape mutations to neutralizing antibodies targeting various domains of the spike protein. We also compared our measurements of the functional effects of mutations to the spike protein to previous studies and natural sequence

data, and found our results were more correlated with enrichment of mutations in natural sequences. This approach can be used to rapidly characterize mutation effects on function and escape from neutralization by sera or therapeutics for emerging pathogens.

We also tailored the lentiviral vector system to map the neutralizing specificity of human anti-HIV sera. We designed mutant libraries of HIV Envelope mutants with combinations of mutations based on previous deep mutational scanning studies and natural sequence data. We used these libraries to map escape from neutralizing antibodies and human sera targeting CD4 binding site of HIV Envelope. Individual mutation effects and antibody epitopes were inferred using a biophysical model to deconvolute mutation effects from the combinations of mutations in the mutant libraries. Most sera mapped had neutralizing specificities similar to individually characterized monoclonal antibodies, but the neutralizing specificity of one serum was best explained by two epitopes within the CD4 binding site. Thus, this approach allows us to map multi-epitope targeting polyclonal immunity and can be used to characterize and evaluate infection or vaccination elicited polyclonal antibody responses to viral entry proteins.

Table of contents

Acknowledgements

I will always feel lucky to have been a member of the Bloom lab, for so many reasons. So, I want to first thank Jesse Bloom. I would describe Jesse as someone who cares about many things; open science, virology, reproducibility, mentorship, using proportional ink on log plots, and more. But if you asked me what Jesse is like as a mentor, then I would say what really matters is that Jesse cares the most about the people in the lab. I have watched Jesse train quite a few amazing scientists now, and his ability to adapt to be the best mentor he can try to be for each person is something I wish I could say was easy to pick up. Thanks Jesse.

I want to thank Julie Overbaugh for co-advising me and the Overbaugh lab for being a great place to learn about HIV. Julie, I will always be thankful for your astounding patience with me procrastinating acting on your recommendations for experiments and graduate school in general only to find they were quite helpful later. I also want to thank the rest of my committee, Neil King, Erick Matsen, Daphne Avgousti, and Adam Geballe. Every conversation I had with each of you felt helpful and I appreciated your genuine interests in the project.

There are too many members of the Bloom lab, past and present, for me to thoroughly thank everyone enough. I want to first thank Hugh, who trained me starting day one in the lab. It was an intimidating place to be, but I knew I would like it when Hugh asked me if I spit in the first PCR I set up for good luck. Thank you Adam for trying the best you could to get me to read more HIV literature, and for being such a huge influence along with Hugh on all the work I did. I want to give a huge thanks to Kate, who worked closely with me during the hardest years of getting the lentivirus system up and running. None of this would have ever been finished if it was not for your hard work and perseverance. Thank you Sarah for being so inviting and encouraging to me and many others in the lab. Thank you Allie for always making time to talk about science or anything else I was going through. Thank you Andrea for keeping the lab running (along with Ariana, Teagan, Keara, and Rachel) and for always talking to me when I

want to go on random science tangents. Thank you Bernadeta for helping the lentivirus system cross the finish line, and thank you to Brendan, Arjun, Caleb, and Tim for the continuing work with the system. Thank you David, Tyler, Frances, Will, and Andrew for making the lab a fun place to do science and sometimes go grab brunch. All of you are amazing scientists and I hope I get to work with you again in the future.

I have been very fortunate to call many people my friends during my PhD. Again, it is too many people to do everyone justice here, but I want to thank everyone in the Spice Party (formerly the Hungry Biker Gang), everyone that I have been fortunate enough to meet through Taylor's work and training, and the steadily increasing number of people Taylor and I have gotten to know through volleyball. You know who you are, and I want you to know that I always look forward to the next time we hang out.

Thank you to my family. Thank you Dylan, Rachel, and Drew. Even though it has been hard living so far from each of you, I cherish all the time we spend together online and when we can see each other in person. Thank you to all my family who have come to visit Taylor and I and have been so supportive of me; Kara, Shelby, Kevin, Mary, Robin, Mimi, and Mackenzie. Mary, thank you for being so excited for me to go to graduate school that you started buying Cal gear because it was the first place at which I interviewed.

Thank you to Eddie and Andrea, my mother and father. Thank you for always being so supportive of me in so many ways. You have always encouraged me to pursue my interests, and you have always pushed me to do the best that I can. I want you to know that I will be forever grateful you did the same as parents to my siblings and me.

Most importantly, thank you Taylor. Six years ago, I asked you to leave home with me without much of a plan. I was still figuring out a lot of things but I already knew at the time that I would be alright with you. What I did not know at the time was how much your compassion and dedication would grow during this time toward so many people and endeavors. I am a lucky person for many reasons, but I am the most thankful that you directed a lot of that toward me.

Chapter I: Introduction

1.1 Viral entry proteins

1.1.1 Introduction to viral entry proteins

Viruses use viral entry proteins on their surfaces to bind and enter host cells. While there are a wide variety of strategies different types of viruses use to accomplish this, this dissertation will focus on viral entry proteins of enveloped viruses that infect humans. These proteins must bind receptors on the cell and then fuse the virus and host cell membranes so that the contents of the virus can enter the cell[1]. There are three main described classes of viral entry proteins for enveloped viruses, each of which have slightly different typical structures[1]. This dissertation will focus on Type I fusion proteins, which are trimers of heterodimers where three trimerized copies of a precursor protein are cleaved into the heterodimers by a protease[1]. After this cleavage, the fusion protein exists in a metastable pre-fusion state and can be irreversibly triggered to accomplish membrane fusion, after which it will be in its stable post-fusion state[1]. Viral entry proteins from different viruses can have different types of triggers for the conformational changes that initiate membrane fusion[1]. For example, influenza Hemagglutinin is triggered by a drop in pH once the virion enters endosomes, HIV Envelope is triggered by simultaneous primary and co-receptor binding, and SARS-CoV-2 spike is triggered by cleavage during endocytosis[1,15].

1.1.2 Immune evasion by viral entry proteins

As critical proteins exposed on the surface of the virus, viral entry proteins are major targets of the host adaptive immune response[2]. Antibodies can bind viral entry proteins and neutralize viruses by blocking receptor binding, locking the viral entry protein into specific conformations, or blocking steps of the membrane fusion process[2]. Viral entry proteins have

evolved several strategies to evade host immunity. Conserved amino acid residues of viral entry proteins, such as those in receptor binding sites, are often tucked away from the outer surface of the protein and masked by immunodominant, mutationally tolerant regions[2–4]. Viral entry proteins are usually decorated with host-derived glycosylations, which are not typically recognized by host immunity and can occlude conserved regions of the protein[3,5,6]. By directing host immunity to mutationally tolerant regions, viruses are able to mutate over time to evade neutralization.

Viruses face significant evolutionary pressure to evade host immunity to their entry proteins. Some viruses which cause self-limiting infections, like influenza and SARS-CoV-2, steadily acquire mutations at a population level over time, resulting in 'ladder-like' phylogenies where more immune-resistant viruses replace older strains[7,8]. HIV causes chronic, life-long infection, which results in continuous within-host evolution and immune escape. As a result, HIV Envelope genes from different strains of HIV can differ by as much as 35% in amino acid identity, making it extremely difficult for antibodies to recognize strains from across the HIV phylogeny[9]. In these ways, each virus presents unique challenges for immunity attempting of recognizing circulating virus strains produced by continuous virus evolution.

A longstanding goal of the field is to develop vaccines which elicit long lasting immunity that is robust in the face of virus evolution. Such immunity does exist for some viruses; antibodies elicited by vaccination against measles virus target many different regions, or epitopes, on its receptor binding protein[10]. This is likely due to several independent antibody lineages targeting the protein, referred to as polyclonal immunity[10,11]. As a result, measles needs to acquire many mutations in a single step to become resistant to this polyclonal immunity, which it is unable to accomplish[10,11]. Unfortunately, this is not the usual case for infection or vaccination-elicited immunity. For example, human serum neutralization of influenza appears similar to a single (or monoclonal) antibody, where only one mutation to its viral entry protein can have a huge effect on escape from neutralization[12]. The following sections go into

more detail on the difficulty of eliciting robust polyclonal immunity for two viruses; SARS-CoV-2 and HIV.

1.1.3 SARS-CoV-2 Spike

The SARS-CoV-2 pandemic began in late 2019 and quickly developed into a major public health crisis, leading to a great deal of interest in studying its viral entry protein, referred to as the SARS-CoV-2 spike protein. SARS-CoV-2 spike is a trimer of heterodimers formed by the cleavage of three S proteins into three S1 and S2 heterodimers[13,14]. It binds the host receptor ACE2, and its fusion process is triggered by an additional cleavage near its fusion peptide during endocytosis[15]. Prior to the SARS-CoV-2 pandemic, the spike proteins of related coronaviruses had been studied, and mutations to stabilize the pre-fusion conformation of these spikes were discovered[16]. The same mutations were used to stabilize SARS-CoV-2 spike to study its structure and were immediately incorporated into vaccines in order to improve neutralizing antibody responses against the virus[17].

After vaccines were approved following phase three trials[18], a major remaining question was whether infection or vaccination would result in long lasting immunity similar to measles, or if antibody resistance would evolve similar to influenza[12,19–21]. As time went on, it became clear that antibody resistance mutations would continue to evolve, and that one or a few mutations could cause a large drop in neutralization from infection of vaccination elicited sera[22–24]. Many high-throughput molecular biology systems (described in section 1.2.1) were used to study this process and try to predict what mutations might arise so they could be monitored in natural sequences. The SARS-CoV-2 pandemic and the rapid development of vaccines using prior knowledge of mutation effects on coronavirus spike structures demonstrated the importance of developing methods for studying viral entry proteins in a safe and high-throughput manner.

1.1.4 HIV Envelope

HIV Envelope (Env) is the conformationally dynamic viral entry protein of HIV. It is a trimer of heterodimers produced by the cleavage of three trimerized gp160 proteins into three gp41 and gp120 heterodimers[25]. In order to initiate fusion, Env first binds its primary receptor, CD4[26]. This process is conformationally dynamic, where Env transitions between a pre-fusion 'closed' conformation, an intermediate 'open' conformation that may be partially bound by CD4, and a CD4-bound conformation[27,28]. After the conformational change induced by CD4 binding, Env exposes its coreceptor binding sites, which can bind CCR5 or CXCR4[28]. Coreceptor binding then initiates the rest of the membrane fusion process[28].

Env is widely considered to be the most difficult target for experimental vaccines due to its complex conformational dynamics, immune evasion tactics, and the requirement for an effective vaccine to result in sterilizing immunity to prevent integration of HIV into the host genome. Through its receptor-dependent conformational shifts, Env prevents conserved residues from being exposed to the immune system for extended periods of time[29]. Env is heavily glycosylated, with an inter-strain median of 25 host-derived N-linked glycosylations present on each heterodimer depending on the strain[5,9]. This glycan shield prevents many regions of the protein from being recognized by the host immune system. Additionally, the 'variable loops' of Env are highly tolerant of insertions, deletions, glycosylation shifts, and mutations[9]. These structures distract immunity from conserved regions and can quickly mutate to evade antibodies targeting them[9,30].

HIV exhibits extensive within-host and population level diversity[7,9,30]. As a retrovirus, HIV reverse transcribes its RNA genome into DNA, integrates into host genomes, and persists over time. The reverse transcription process is error prone, so the virus generates a high level of diversity within each host[31,32]. Neutralizing antibodies do develop against replicating viruses, but these antibodies are usually quickly escaped by the virus population[33,34]. Rarely, antibodies

develop that are capable of neutralizing not only the within host viruses they coevolved with, but many strains of HIV from across its global phylogeny[4,35–37]. Many of these 'broadly neutralizing antibodies' have been described, which target a variety of different conserved epitopes on Env[36]. The structures and evolutionary trajectories of broadly neutralizing antibodies have informed experimental vaccine designs that attempt to recapitulate their development through vaccination[4,35,38]. However, experimental vaccines still have great difficulty in eliciting effective broadly neutralizing antibodies in animal models against single epitopes of Env, let alone the multiple epitopes likely required for polyclonal immunity robust to virus evolution[39–41].

1.2 High-throughput methods for characterizing mutation effects for viral entry proteins

1.2.1 Overview of previous methods

High-throughput methods for characterizing mutation effects have transformed how mutations to many proteins and sequence elements are studied, including viral entry proteins[42,43]. In prior low-throughput methods, mutations of interest are individually cloned into viral entry proteins and used in replicative full-length genome virus assays or in pseudovirus assays. Often these viruses are incubated with dilutions of antibodies or serum in neutralization assays to determine the degree to which they are neutralized at different concentrations of antibody or serum. While informative, this is time and resource consuming, and can be limited due to biosafety concerns. High-throughput methods for characterizing mutation effects, often referred to as deep mutational scanning, use pools of mutants containing many or all possible amino acid mutants to a protein of interest. This 'mutant library' pool is used in selections such as protein function assays or antibody neutralization assays where non-functional mutants or antibody-neutralized mutants are depleted. Deep sequencing is then used to measure which

mutants were enriched and which mutants were depleted, revealing which mutations are either beneficial or deleterious in the selection used.

Many high-throughput techniques have been developed to study viral entry proteins[43,44]. Deep mutational scanning techniques have been used with replicative full-length genome viruses to simultaneously measure the effects of all mutations to viral entry proteins for some viruses[45–51,12]. However, previous described deep mutational scanning of viral entry proteins has been unable to measure the effects of combinations of mutations across entire proteins and is still limited to few viruses due to biosafety concerns. Phage display techniques can allow screening of antibody binding against mutant libraries of several virus strains or even different viruses at once, but is limited to short, linear epitopes and binding assays rather than virus neutralization[52–56]. Yeast display can be used to display mutant libraries of larger domains of viral entry proteins, but is still limited to antibody or soluble receptor binding assays[57–60]. Cell surface display can display mutant libraries of full viral entry proteins, but antibody binding or cell-fusion inhibition assays often rely on cell sorting methods that can limit the throughput of these experiments and therefore the number of mutations and sites assayed[61–63]. Another relevant method for studying viral entry proteins is electron microscopy-based polyclonal epitope mapping (emPEM). emPEM can be used to visualize how and where antibodies in polyclonal serum bind to viral entry proteins[64–66], but like other techniques listed above, this measures binding by antibodies rather than virus neutralization.

1.2.2 Introduction to lentiviral vectors and their use for deep mutational scanning

Lentiviral vectors have been proven a useful tool for studying viral entry proteins. Lentiviral vectors are pseudoviruses typically derived from HIV that are only capable of one round of replication[67]. Lentiviral vector pseudoviruses are produced from a few main parts. The first is the lentiviral genome or backbone, which is a shortened version of the HIV genome with most of the HIV gene sequences removed and replaced with reporters for infection such as

fluorescent proteins or Luciferase. The second is the lentiviral helper plasmids, a separate set of plasmids that express genes necessary to form the virions and complete one round of replication and integration into cells, such as Gag-Pol, Tat, and Rev. The final part is a viral entry protein. Importantly, lentiviruses can be 'pseudotyped' with a wide variety of viral entry proteins by expressing any amenable viral entry protein alongside the lentiviral helper plasmids[68,69]. The resulting virions will be coated with the heterologous viral entry protein and rely on it for entry into cells. In this way, viral entry proteins from emerging pathogens or pathogens without reverse genetics systems can be studied. Pseudotyping of lentiviruses with SARS-CoV-2 spike was quickly developed as a pseudovirus system to study spike near the beginning of the SARS-CoV-2 pandemic and became widely used for spike neutralization assays[70].

In the following chapters, I describe our use of a lentiviral vector to perform deep mutational scanning. This has several advantages over viral entry protein characterization methods described above. Since lentiviral vectors rely on the viral entry protein they are pseudotyped with to enter cells, they can be used to measure how mutations affect both viral entry protein function of entering cells and neutralization by antibodies or serum, rather than just displaying a portion of the protein or only measuring antibody binding. By inserting viral entry protein mutants into the lentivirus genome and using a two-step protocol for generating the viruses (See Figure 2.2 in Chapter II), we are able to establish a genotype-phenotype link where a mutant gene sequence within the genome matches the viral entry protein on the surface of the virion. In order to easily measure the effect of many mutations and combinations of mutations simultaneously, we include a unique random-nucleotide barcode with each mutant. We are able to add these barcodes to mutants because there is less restriction on added sequences to lentiviral vector genomes compared to the replicative full-length genome viruses used for prior deep mutational scanning, and our ability to control the lentivirus replication cycles allows us to prevent the mutant genes and barcodes from becoming scrambled due to recombination during

virus replication cycles[71–73]. After first linking which mutants are with which barcodes using long read PacBio sequencing, downstream experiments only need short read sequencing of the barcodes to determine which mutants were selected for in functional assays or neutralization assays.

The development of the lentiviral vector deep mutational scanning system was a collaborative project between Katharine (Kate) Crawford, Bernadeta Dadonaite, and myself. I, along with Kate and Bernadeta, engineered and optimized the lentiviral genome for deep mutational scanning, engineered and tested cell lines for their use with the system, and optimized virus culture, concentration, infection, and sequencing conditions. I applied the system to deep mutational scanning of the HIV Envelope protein, while Bernadeta applied the system to SARS-CoV-2 spike and Kate applied the system to LASV-GP (unpublished). Because the development and application of this system to each viral entry protein were integrated at most steps, I have included the SARS-CoV-2 experiments and results in this thesis despite not having personally performed the SARS-CoV-2 deep mutational scanning experiments. This is also reflected by the co-first authorship of the version of Chapter II published in *Cell.* The application of this system to SARS-CoV-2 spike is included in this thesis because it is an important demonstration of the utility and adaptability of the lentivirus deep mutational scanning system.

1.2.3 Methods for deconvoluting mutation effects from mutants with combinations of mutations

Measuring or predicting the effects of combinations of mutations is necessary to map escape to polyclonal immunity targeting multiple epitopes at once, but the effects of combinations of mutations can be hard to predict due to epistasis[74,75]. Epistasis is non-additivity in the effects of mutations. It can be specific, where residues interact directly or allosterically to have unpredictable effects when mutated together, or it can be global, where some non-

additivity in mutation effects can be explained by all mutations having additive effects on a non-measured latent scale with a non-linear relationship to the observed scale of mutation effects[74]. While specific epistasis can be almost impossible to predict, methods for modeling global epistasis have been developed[76,77]. These global epistasis models attempt to model each mutation's effects on the latent additive scale and the function used to transform this scale to the observed scale of mutation effects. In this way, individual mutation effects can be deconvolved from libraries of mutants with combinations of mutations. We also recently described a biophysical model that uses a similar approach to deconvolute the effects of mutations on antibody neutralization at multiple epitopes on viral entry proteins[78]. Under this model, antibody neutralization at each epitope depends on the antibody or serum concentration, and mutations within a given epitope have additive effects on antibody affinity. The overall neutralization of any particular mutant then depends on the antibody or serum concentration, how much the mutations the mutant has escape the neutralization at each epitope, and how much the neutralization activity at each epitope contributes to overall neutralization. Together with the lentiviral vector deep mutational scanning system, we use these methods to measure and predict the effects of mutations on viral entry protein function and escape from polyclonal immunity in the following chapters.

Chapter II: A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike

A version of this chapter has been published as:

*Dadonaite, B., *Crawford, K. H., *Radford, C. E., Farrell, A. G., Timothy, C. Y., Hannon, W. W., ... & Bloom, J. D. (2023). A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike. *Cell*, *186*(6), 1263-1278.

*these authors contributed equally

As stated in section 1.2.2, this chapter describes the collaborative development of the lentivirus deep mutational scanning system by Bernadeta Dadonaite, Kate Crawford, and myself. Although I did personally perform the SARS-CoV-2 spike deep mutational scanning experiments, these experiments are included in their entirety in this chapter with the development of the lentivirus system because the development and experiments were integrated with each other and the SARS-CoV-2 experiments demonstrate the utility and adaptability of the lentivirus deep mutational scanning system.

2.1 Abstract

A major challenge in understanding SARS-CoV-2 evolution is interpreting the antigenic and functional effects of emerging mutations in the viral spike protein. Here we describe a new deep mutational scanning platform based on non-replicative pseudotyped lentiviruses that directly quantifies how large numbers of spike mutations impact antibody neutralization and pseudovirus infection. We demonstrate this new platform by making libraries of the Omicron BA.1 and Delta spikes. These libraries each contain ~7000 distinct amino-acid mutations in the context of up to ~135,000 unique mutation combinations. We use these libraries to map escape

mutations from neutralizing antibodies targeting the receptor binding domain, N-terminal domain, and S2 subunit of spike. Overall, this work establishes a high-throughput and safe approach to measure how ~105 combinations of mutations affect antibody neutralization and spike-mediated infection. Notably, the platform described here can be extended to the entry proteins of many other viruses.

2.2 Introduction

The spike protein is the key target of neutralizing antibodies against SARS-CoV-2. Unfortunately, spike has undergone rapid evolution which has eroded the potency of serum neutralization and enabled escape from most monoclonal antibodies[79–82]. Deep mutational scanning experiments can prospectively measure the effects of large numbers of mutations even before they emerge in viral variants, and therefore have been a valuable tool for rapidly interpreting how newly observed mutations in the spike affect antibody binding and protein folding or function[58,83,84]. The high-throughput nature of deep mutational scanning experiments has also enabled the generation of huge datasets that can inform computational methods for predicting the antigenic properties of possible future viral variants[83,85].

However, prior deep mutational scanning of the SARS-CoV-2 spike has been limited to either solely focusing on the receptor-binding domain (RBD)[24,57,83], other subdomains[63,86] or just a small number of mutations across spike[87]. Furthermore, all previous spike deep mutational scanning experiments have been based on cell-surface display using either yeast[57,88] or mammalian cells[63,86,87], and therefore are limited to measuring antibody binding rather than neutralization, despite the fact that neutralization is thought to be a more relevant correlate of protection[89,90].

Here we describe a new deep mutational scanning platform that directly measures how mutations affect cellular infection and antibody neutralization in the context of the full SARS-

CoV-2 spike pseudotyped on non-replicative lentiviral particles. The key innovation behind the platform is a two-step pseudovirus generation protocol that enables creation of large pseudovirus libraries with a link between the lentiviral genotype and the particular spike protein variant on the pseudovirus surface. We demonstrate that this new platform can be used to create large genotype-phenotype linked pseudovirus libraries and map how mutations to spike affect both cellular infection and neutralization by antibodies targeting diverse regions of spike, including the RBD, N-terminal domain (NTD), and S2 subunit.

2.3 Results

2.3.1 Producing pseudoviruses with genotype-phenotype link

To characterize thousands of mutations in spike glycoprotein, we first established a lentiviral pseudotyping platform that maintains a genotype-phenotype link between the lentiviral genome and the spike variant on the virion's surface. Lentiviral spike-pseudotyping usually involves transfection of a backbone that carries a reporter gene flanked by the lentiviral long terminal repeats (LTRs), helper plasmids that code for structural and nonstructural genes required for the lentiviral life cycle, and an expression plasmid that codes for the spike variant of interest (Figure 2.1A)[68,70,91]. When these components are transfected into producer cells, virions are formed that carry lentiviral genomes and display spikes on their surface. However, because genome incorporation into a virion does not depend on the expressed spike, there is no link between the virion's genotype and the phenotype of the spike on its surface (Figure 2.1A). The absence of a genotype-phenotype link is not problematic when only a single spike variant is used for transfection—however, it precludes deep mutational scanning studies that involve studying thousands of variants in a single pooled experiment.

To create a lentiviral genotype-phenotype link, we first generated a lentivirus backbone with the following key elements (Figure 2.2A): (1) we restored the ability of the lentivirus to

transcribe its full genome after integration by repairing the 3' LTR deletion present in traditional lentivirus vectors[67,92], (2) we placed spike in the lentivirus backbone under an inducible promoter, (3) we added a second constitutive promoter driving both a fluorescent reporter (ZsGreen) and a puromycin resistance gene.

Figure 2.1 Pseudovirus titers from phenotype-genotype linked lentiviruses

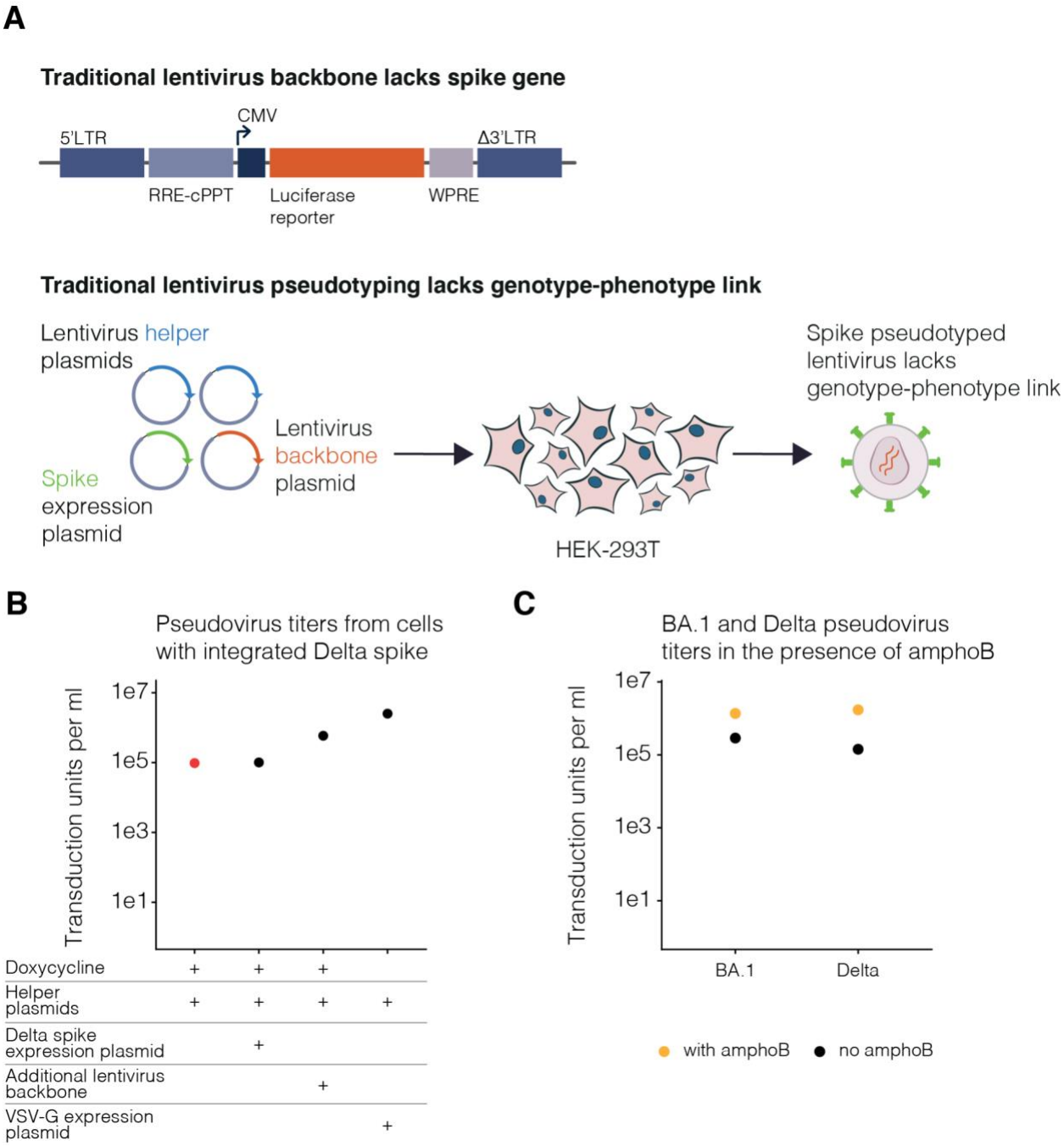


Figure 2.1: (A) Traditional lentivirus pseudotyping method. Lentivirus backbone used for pseudotyping does not code for spike gene. To make spike pseudotyped lentivirus, lentiviral helper plasmids, backbone and spike expression plasmid are transfected into producer cells to make spike pseudotyped lentivirus. This method produces lentiviruses that lack genotype-phenotype link because the spike expressed on the surface of a viral particle is not coded by the lentiviral genome. (B) Delta spike pseudotyped lentivirus titers. Viruses were produced under indicated conditions from cells with integrated lentivirus genomes carrying Delta spike. Virus titers for conditions used to generate the actual deep mutational scanning libraries are coloured

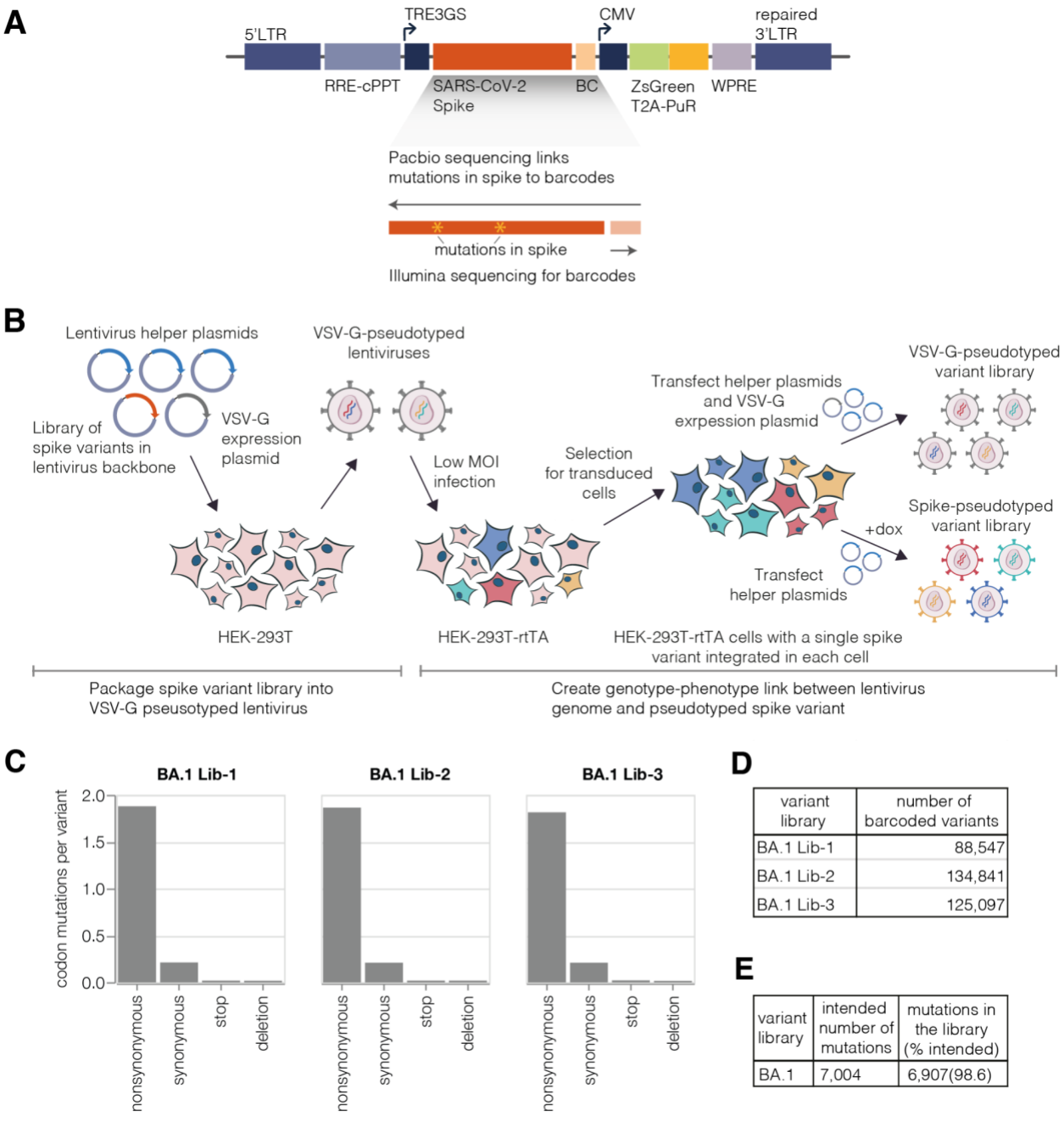
red. Viruses were titrated on ACE2-TMPRSS2-HEK-293T cells. (C) BA.1 or Delta spike-pseudotyped lentivirus titers in the presence or absence of amphotericin B (amphoB). BA.1 virus was titrated on ACE2-HEK-293T cells and Delta virus was titrated on ACE2-TMPRSS2-HEK-293T cells.

Next, we developed a multi-step protocol that creates a genotype-phenotype link by ensuring that each producer cell only expresses a single variant of spike (Figure 2.2B). In the first step of this protocol, we transfect cells with the spike-encoding backbone, a VSV-G expression plasmid, and the necessary helper plasmids. This produces non-genotype-phenotype-linked VSV-G-pseudotyped lentiviruses that we use to infect target cells at low multiplicity of infection, so that most infected cells receive no more than one lentiviral genome. Next we select for cells with integrated lentiviral genomes using puromycin, which yields a population of cells where each cell stores only a single spike variant. The spike is under an inducible promoter, which is only activated by addition of doxycycline. To produce virions, we induce spike expression with doxycycline and transfect the helper plasmids necessary to produce lentiviruses. We validated that this approach can be used to generate genotype-phenotype linked spike-pseudotyped viruses with titers >105 transduction units per ml (Figure 2.1B). We can further increase viral titers by ~5-10 fold by infecting cells in the presence of amphotericin B, as has been reported previously[93–95].

2.3.2 Design of mutations in SARS-CoV-2 spike deep mutational scanning library

Rather than create deep mutational scanning libraries containing all possible amino-acid mutants of spike, we chose to introduce only mutations that seem likely to arise during natural evolution and yield a functional spike protein. We had two rationales for designing our libraries in this way: (1) it reduces the total number of mutations that need to be included in the library, and (2) it increases the probability that variants with multiple mutations will remain functional by reducing the fraction of mutations that are highly deleterious.

Figure 2.2 Deep mutational scanning platform for spike



variant library	number of barcoded variants
BA.1 Lib-1	88,547
BA.1 Lib-2	134,841
BA.1 Lib-3	125,097

variant library	intended number of mutations	mutations in the library (% intended)
BA.1	7,004	6,907(98.6)

Figure 2.2: (A) Lentivirus backbone used for deep mutational scanning. The backbone contains functional lentiviral 5′ and 3′ long terminal repeat (LTR) regions. The spike gene is under an inducible tet response element 3rd generation (TRE3G) promoter, and there is a 16 nucleotide barcode (BC) downstream of the stop codon. A CMV promoter drives expression of reporter ZsGreen gene that is linked to a puromycin resistance gene (PuR) via a T2A linker. The backbone also contains a woodchuck hepatitis virus posttranscriptional regulatory element (WPRE), Rev response element (RRE), and a central polypurine tract (cPPT). (B) Approach for creating genotype-phenotype linked lentivirus libraries. HEK-293T cells are transfected with

spike-carrying lentivirus backbone, VSV-G expression plasmid and lentiviral helper plasmids to generate VSV-G-pseudotyped lentiviruses. These viruses are used to transduce reverse tetracycline-controlled transactivator (rtTA) expressing HEK-293T cells at low multiplicity of infection (MOI), and successfully transduced cells are selected using puromycin. Selected cells can be transfected with helper plasmids and a VSV-G expression plasmid to produce VSV-G-pseudotyped viruses carrying all genomes present in the deep mutational library or selected cells can be induced with doxycycline (dox) to express spike and transfected with only the helper plasmids to generate spike-pseudotyped lentiviruses that have a genotype-phenotype link. (C) Average number of mutations per barcoded spike in BA.1 libraries. (D) Total number of barcoded variants in each BA.1 library. (E) The coverage of intended mutations across all BA.1 libraries.

Specifically, we included only mutations that have been observed in spike sequences deposited on the GISAID database[96], reasoning that these mutations would represent mostly functional spike proteins. We introduced mutations at a higher frequency when they have emerged in spike independently many times according to the pre-built SARS-CoV-2 phylogenies from UShER[97]. Finally, we included every possible amino acid change at sites in spike that are evolving under positive selection[98]. We also included deletions at sites where such mutations are observed frequently in natural SARS-CoV-2 evolution. In total, our library design targeted 7,004 mutations in the BA.1 spike and 6,852 mutations in the Delta spike.

To introduce these mutations in the spike gene we used a PCR-based mutagenesis method with a primer pool containing the desired mutations[45]. Importantly, this method introduces multiple mutations in each spike variant: we targeted ~2 to 3 codon mutations per variant, ensuring the effects of most mutations are measured in multiple genetic backgrounds. The mutated spike genes were then barcoded with 16 random nucleotides placed downstream of the spike-coding sequence (Figure 2.2A), and cloned into the lentivirus backbone. As described below, after integration of the libraries into cells, these barcodes can be linked to the full set of mutations in each spike variant to facilitate downstream sequencing[99,100].

2.3.3 Production of pseudotyped BA.1 and Delta spike deep mutational scanning libraries

We used the genotype-phenotype linked pseudovirus production strategy in Figure 2.2B to make BA.1 and Delta deep mutational scanning libraries. We created three independent BA.1 libraries each containing ~100,000 barcoded variants and two independent Delta libraries each containing ~50,000 barcoded variants (Figures 2.2E, 2.3A; a "barcoded variant" is a spike with a unique nucleotide barcode and some random mutation set; different barcoded variants usually but not always contain different mutations). After integrating the libraries into cells at low multiplicity of infection, we generated VSV-G-pseudotyped lentivirus from these cells by co-transfecting a plasmid expressing VSV-G alongside the other lentiviral helper plasmids (Figure 2.2B, top right). The use of VSV-G-pseudotyped virus ensures that we generate infectious lentiviral virions from all integrated backbones regardless of whether they encode a functional spike mutant. We then infected this VSV-G-pseudotyped lentivirus into a new round of cells, and performed long-read PacBio sequencing to link the barcodes to the full set of spike mutations for each variant. We performed the PacBio barcode-mutation linking after integration into cells because recombination of the pseudodiploid lentiviral genome during integration[72,73] means the barcode-mutation pairings may be different in the integrated cells to those in the original lentiviral backbone plasmids[71]. Importantly, linking barcodes to spike variants allows us to use short-read Illumina sequencing of the barcode to obtain the full spike genotype in all subsequent experiments.

Overall, the sequencing revealed that we had successfully introduced ~99% of the targeted mutations in the BA.1 and Delta spike libraries (Figures 2.2E, 2.3B), with at least one amino-acid mutation at each site in spike for both the BA.1 and Delta libraries. The barcoded variants in the BA.1 libraries had on average ~2 codon mutations per spike, while the variants in the Delta libraries had ~3 codon mutations per spike (Figures 2.2C, 2.3C). The number of

mutations per variant is roughly Poisson distributed, so some variants had zero or one mutation, while others had many more (Figure 2.4).

Figure 2.3 Characteristics of Delta spike deep mutational scanning libraries and BA.1 library functional scores

A

variant library	*number of barcoded variants*
Delta Lib-1	49,053
Delta Lib-2	48,855

B

variant library	*intended number of mutations*	*mutations in the library(% intended)*
Delta	6,852	6,771(98.8)

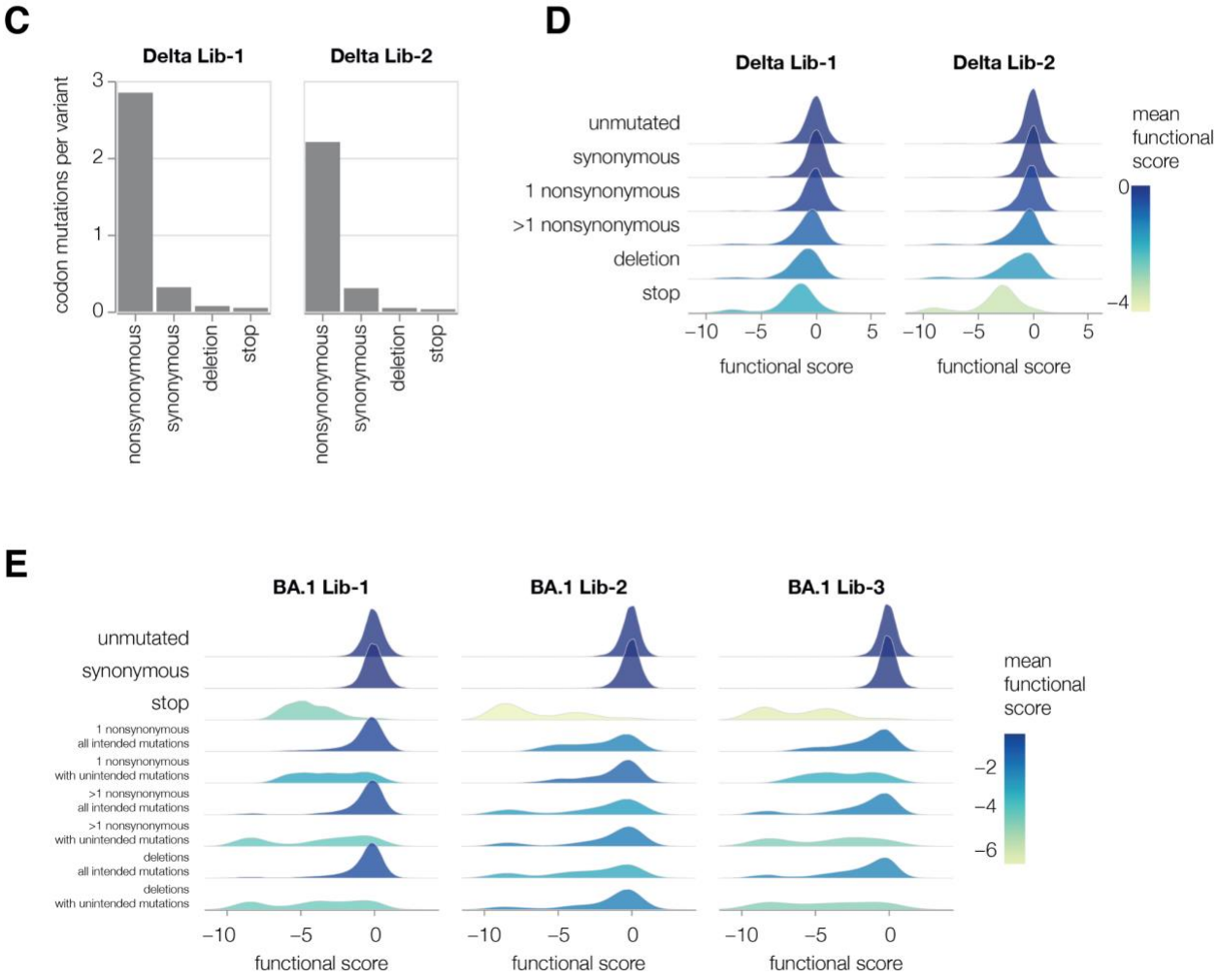
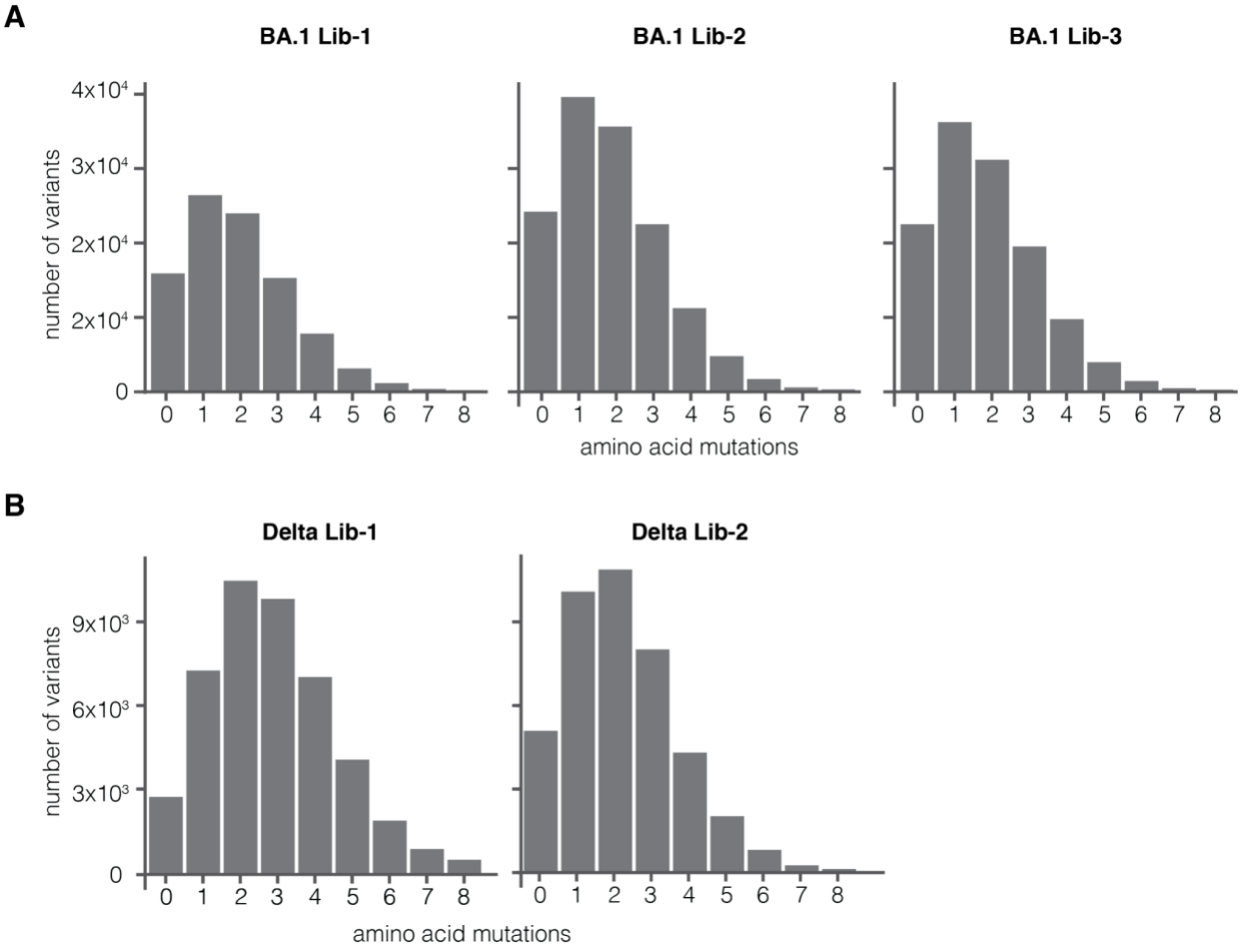


Figure 2.3: (A) Total number of barcoded variants in each Delta library. (B) Coverage of intended mutations across both Delta libraries. (C) Average number of mutations per barcoded spike in Delta libraries. (D) Distribution of functional scores for variants with different types of mutations in the Delta libraries. (D) Distribution of functional scores for variants in BA.1 libraries

stratified by whether they only contained mutations “intended” during the library design or also contain other non-intended mutations.

Figure 2.4 Distribution of the number of amino-acid mutations per variant in BA.1 (A) and Delta (B) deep mutational scanning libraries



We then generated the actual spike-pseudotyped deep mutational scanning libraries from the variants stored at single copy in the cells (Figure 2.2B, lower right). We calculated a functional score for each variant based on its relative frequency in the spike- versus VSV-G-pseudotyped libraries. Positive functional scores indicate spike variants mediate pseudovirus infection better than the parental spike, whereas negative functional scores indicate worse pseudovirus infection. As expected, spike variants with premature stop codons had highly

negative functional scores, while unmutated and synonymously mutated spike variants had functional scores close to zero (Figures 2.3D-E, 2.5). Some variants with nonsynonymous mutations had functional scores close to zero, while others had more negative scores, reflecting the fact that some but not all nonsynonymous mutations are deleterious (Figures S2.3D, 2.5; recall that our library design protocol preferentially introduced nonsynonymous mutations expected to yield functional spikes). Variants with multiple nonsynonymous mutations tended to have lower functional scores than variants with just one nonsynonymous mutation (Figures 2.3D, 2.5), reflecting the cost of accumulating multiple often mildly deleterious mutations. Variants that only contained mutations designed into the libraries tended to have higher functional scores than variants that also carried other unintended nonsynonymous mutations (Figure 2.3E), suggesting that our library-design strategy indeed enriched for functionally tolerated mutations.

Figure 2.5 Some mutations tend to impair spike-mediated pseudovirus infection

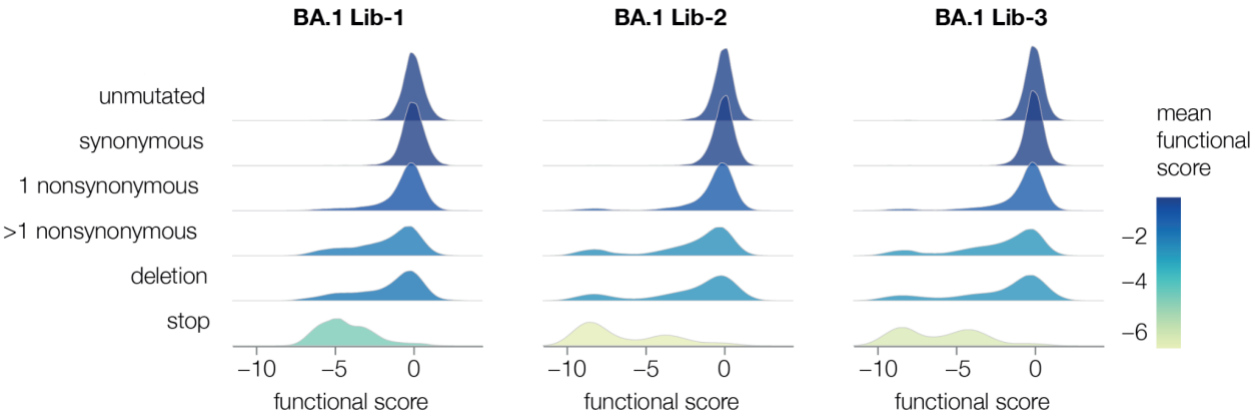


Figure 2.5: For each barcoded spike variant, we compute a functional score that reflects how well that spike mediates pseudovirus infection relative to the unmutated spike: negative scores indicate impaired infection, positive scores indicate improved infection. The plots show the distribution of functional scores across all variants in each of the three BA.1 libraries for different categories of variants, with each distribution colored by the mean functional score for that variant type.

2.3.4 Use of an absolute standard to measure viral neutralization by deep sequencing

Traditional neutralization assays measure the infectivity of a single virus variant at multiple antibody concentrations. Deep sequencing can measure the relative infectivities of many viral variants in pooled infections in the presence of an antibody. However, to convert the relative infectivities measured by deep sequencing into actual neutralization values, it is necessary to have an absolute standard that does not vary in its infectivity as a function of antibody concentration (Figure 2.6A). To enable such measurements in our experiments, we added a barcoded VSV-G-pseudotyped virus into our libraries. This virus is produced from a separate cell line that has integrated lentiviral backbones carrying a barcoded mCherry gene (see Methods). Importantly, this VSV-G-pseudotyped virus is not neutralized by any of the spike-binding antibodies (Figures 2.6A, 2.7A), and so the counts for VSV-G barcodes provide an absolute neutralization standard. To calculate the non-neutralized fraction for each viral variant, we simply compute the change in its barcode frequencies relative to the VSV-G standard (Figure 2.6B).

To validate this approach, we added the VSV-G absolute standard at ~1% of our BA.1 library titers and incubated the virus library with increasing concentrations of the LY-CoV1404 antibody, as schematized in Figure 2.6B. We then infected the library into ACE2-expressing target cells overnight, recovered viral genomes, and quantified the abundance of each viral barcode using deep sequencing. As expected, the fraction of VSV-G standard reads increased with antibody concentration because fewer spike variants could still infect in the presence of antibody (Figure 2.6C). We then calculated the non-neutralized fraction for each viral variant in our libraries after selection at different concentrations of the antibody. As expected, increasing antibody concentrations led to decreased non-neutralized fraction averaged over variants (Figure 2.6D). Notably, variants with greater number of substitutions had higher non-neutralized fractions, as expected if some substitutions escape the antibody.

Figure 2.6: A VSV-G standard enables measurement of absolute neutralization by deep sequencing

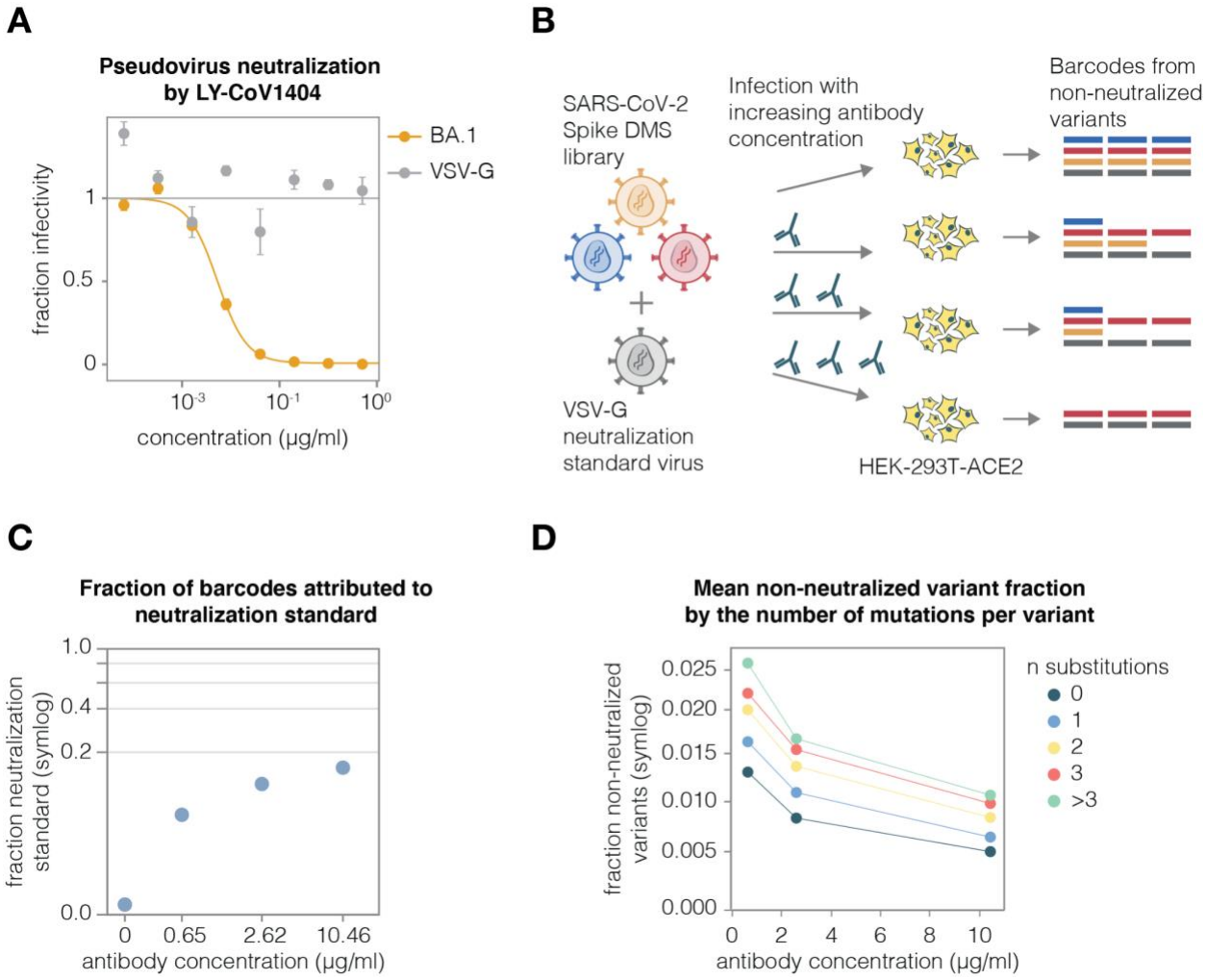


Figure 2.6: (A) Neutralization assay demonstrating that BA.1-spike-pseudotyped lentivirus is neutralized by antibody LY-CoV1404, but the VSV-G-pseudotyped neutralization standard is not. (B) Use of the VSV-G standard to measure absolute neutralization. Deep mutational scanning libraries are mixed with VSV-G neutralization standard. The virus mixture is incubated with a no-antibody control or increasing antibody concentrations and infected into ACE2-expressing 293T cells. After ~12 hours viral genomes are recovered, barcodes are sequenced, and absolute neutralization of each variant is computed by comparing its barcode counts to those from the VSV-G standard. (C) Fraction of barcodes derived from the VSV-G neutralization standard in infections with increasing LY-CoV1404 concentrations. (D) BA.1 deep mutational scanning library non-neutralized fractions averaged across variants with different numbers of amino-acid mutations at differentLY-CoV1404 concentrations.Note panels C and D use a symlog scale.

Figure 2.7 The VSV-G neutralization standard is not neutralized by antibodies 5-7, CC9.104 and CC65.105.

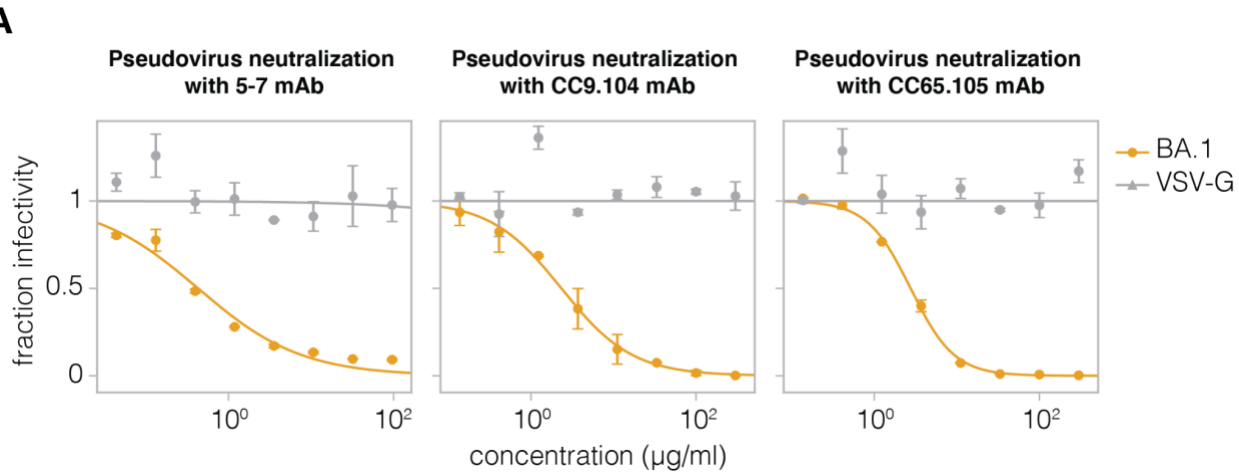


Figure 2.7: (A) Neutralization assays using NTD-targeting 5-7 mAb and S2-targeting CC9.104 and CC65.105 antibodies against lentivirus pseudotyped with BA.1 spike or VSV-G.

2.3.5 Mapping antibody escape using a full spike deep mutational scanning system

To demonstrate that pseudovirus-based deep mutational scanning can map escape from neutralizing antibodies targeting any region of spike, we chose a set of BA.1-neutralizing antibodies that bind distinct regions of spike: RBD-binding LY-CoV1404, NTD-binding 5-7, and S2-binding CC67.105[101–103]. Note that LY-CoV1404, also known as bebtelovimab, is one of the few clinically approved antibodies that retains potency against BA.1, BA.2 and other major Omicron lineages[81,102].

We first mapped escape from LY-CoV1404, applying the approach in Figure 2.6B to our three independent BA.1 libraries, and performing a technical replicate for one library. We used a biophysical model to decompose the measurements for spike variants in our libraries (some of which are multiply mutated) into escape scores for individual mutations[78]. These mutation escape scores correlated well among technical and biological replicates (Figure 2.8A).

Figure 2.8 Antibody LY-CoV1404 escape mapping

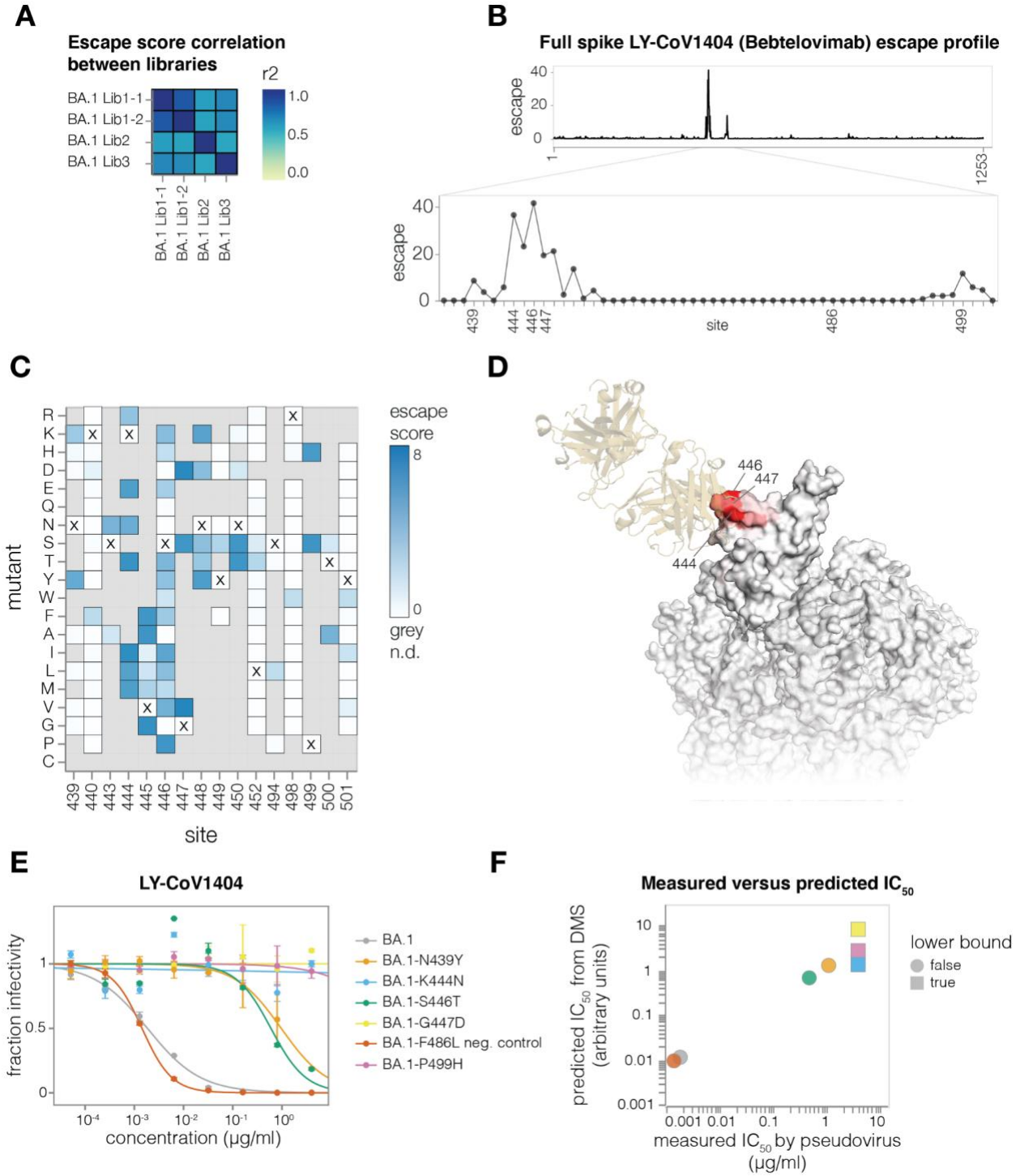


Figure 2.8: (A) Correlation of mutation escape scores between technical replicates (BA1 Lib-1.1 and BA1 Lib-1.2) and biological replicates (BA1 Lib-1, BA1 Lib-2, BA1 Lib-3). (B) Total escape scores at each site in the BA.1 spike, and a zoomed-in plot showing the key escape sites. Sites of mutations chosen for validation experiments are labeled on the x-axis. (C) Heatmap of mutation escape scores at key sites. Residues marked with X are the wild-type amino acids in

BA.1. Amino acids not present in our libraries are shown in gray. An interactive heatmap for the entirety of spike is at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/LyCoV-1404_escape_plot.html (D) Surface representation of spike coloured by sum of escape scores at that site. LY-CoV1404 antibody is in yellow. Only the antibody-bound protomer is coloured. PDB IDs 7MMO and 6XM4 were aligned to make this structure. (E) Validation pseudovirus neutralization assays of the indicated BA.1 spike mutants against the LY-CoV1404 antibody. (F) Correlation between predicted IC50 values from deep mutational scanning (DMS) data versus the IC50 values measured in the validation assays in panel (E). The points are colored as in panel (E). Lower bound indicates that the antibody did not neutralize at the highest concentration tested in the validation neutralization assay. Site numbering in all plots is based on the Wuhan-Hu-1 sequence.To show that we can map escape from non-RBD-targeting antibodies, we next mapped the NTD-targeting 5-7 antibody[101]. This antibody targets an epitope outside the defined antigenic supersite in NTD and is one of the few NTD-targeting antibodies isolated pre-Omicron that still retains some potency against Omicron variants[80,101,104].

As expected, the key LY-CoV1404 escape sites were in the antibody's previously described epitope in the RBD[102], which spans sites 439-452 and 498-501 (Figure 2.8B-D and https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/LyCoV-1404_escape_plot.html). However, our deep mutational scanning emphasizes that only some mutations at these sites escape LY-CoV1404 neutralization. For instance, many amino acid mutations at site 446 strongly escape LY-CoV1404, but mutating this site from G (the identity in Wuhan-Hu-1) to S (the identity in BA.1 and BA.2.75) does not have a large effect. This observation emphasizes the somewhat serendipitous nature of the preserved potency of LY-CoV1404. However, this antibody may soon be escaped because sub-variants of BA.5 and BA.2.75 with mutations in the key escape site of K444 are increasingly being detected[83,105].

To validate the LY-CoV1404 deep mutational scanning, we cloned a set of mutations in the BA.1 spike with a range of effects in the deep mutational scanning data and performed standard pseudovirus neutralization assays (Figure 2.8E). All the tested mutations exhibited neutralization phenotypes consistent with those measured in the deep mutational scanning. Furthermore, the neutralization assay IC50 values correlated well with those predicted by our biophysical model[78] parameterized by the deep mutational scanning data (Figure 2.8F).

Figure 2.9 Comparison between antibody escape mapping using full spike pseudovirus deep mutational scanning versus our previously described yeast-display deep mutational scanning of just the RBD.

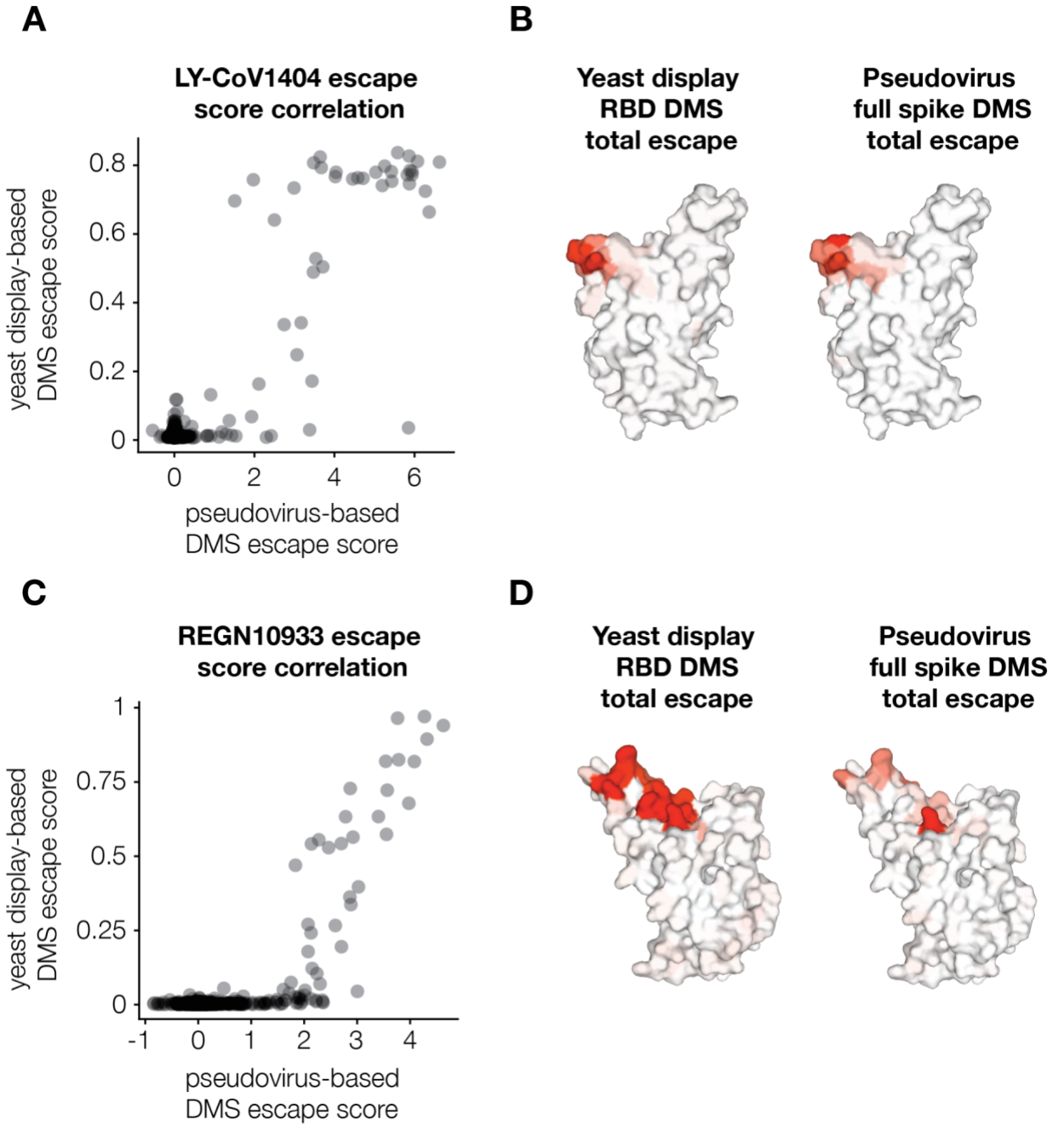


Figure 2.9: (A) Correlation between measured mutation-level escape scores for LY-CoV1404 antibody in pseudovirus and yeast display deep mutational scanning experiments. Yeast display data is taken from[106]. (B) Surface representation of SARS-CoV-2 RBD coloured by sum of escape scores for LY-CoV1404 antibody at that site. (C) Correlation between measured

mutation-level escape scores for REGN10933 antibody in pseudovirus and yeast display deep mutational scanning experiments. Yeast display data is taken from[84] (D) Surface representation of SARS-CoV-2 RBD coloured by sum of escape scores for REGN10933 at that site. PDB ID: 6XM4.

We also compared the full spike LY-CoV1404 deep mutational scanning measurements to results from our previously described yeast-display system for deep mutational scanning of only the RBD[57,88]. The escape scores between the two experimental approaches correlated well (Figure 2.9A) and both methods identified the same epitope (Figure 2.9B).

To show that we can map escape from non-RBD-targeting antibodies, we next mapped the NTD-targeting 5-7 antibody[101]. This antibody targets an epitope outside the defined antigenic supersite in NTD and is one of the few NTD-targeting antibodies isolated pre-Omicron that still retains some potency against Omicron variants[80,101,104].

The deep mutational scanning showed that the key escape sites for 5-7 were in a hydrophobic pocket next to the N4 loop (site 172-178) (Figure 2.10A-C and https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/NTD_5-7_escape_plot.html), consistent with prior structural characterization of this antibody's epitope[101]. In addition, deletions in 167-171 ß-sheet, as well as mutations at the base of the adjacent loops such as G103 and V126 also escaped antibody 5-7 (Figure 2.10A-B). We validated these deep mutational scanning results by performing individual neutralization assays with pseudoviruses containing L176K, S172N and G103F mutations (Figure 2.10D), all of which had the expected effect of completely escaping neutralization (Figure 2.10E).

Figure 2.10 Antibody 5-7 escape mapping

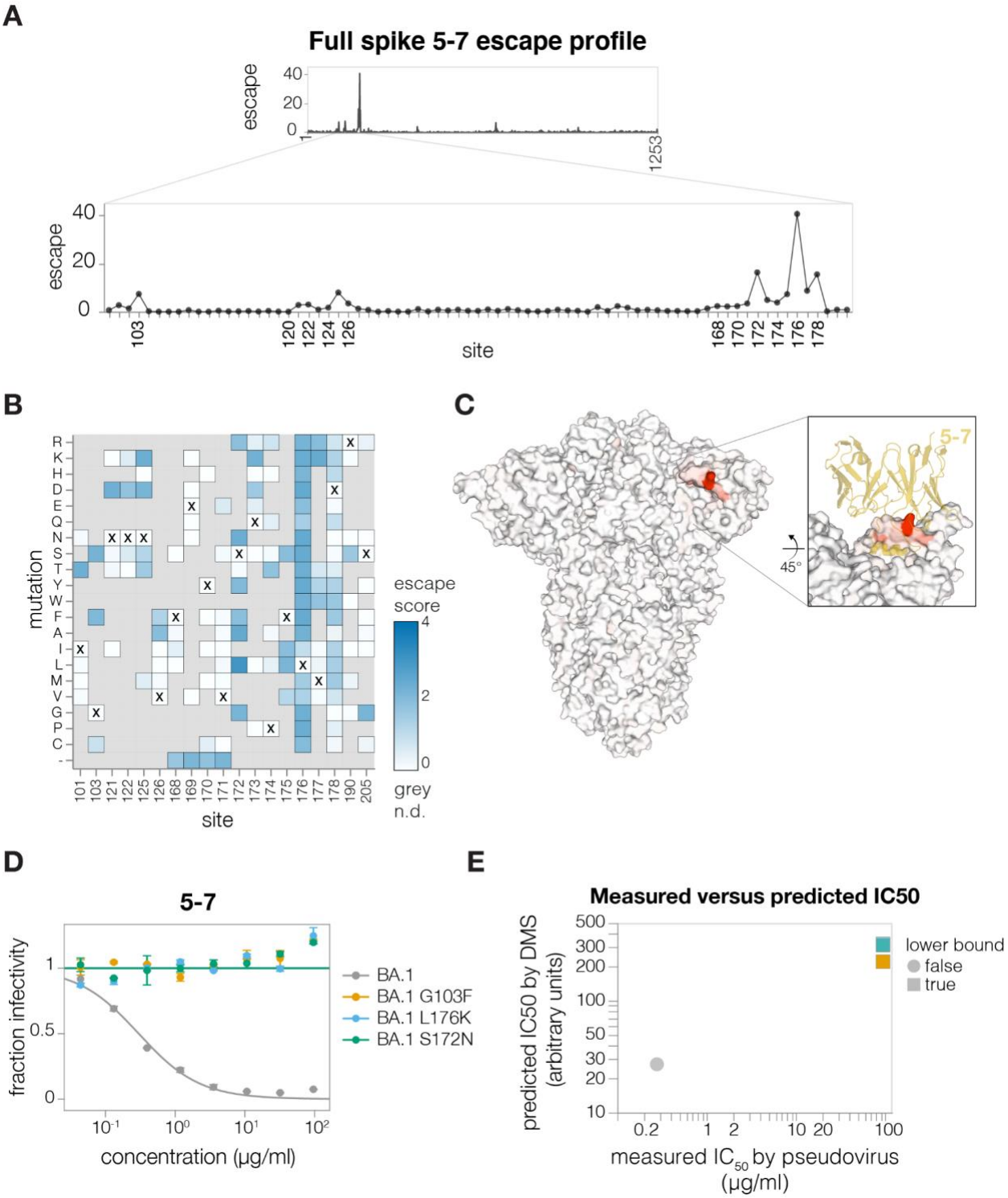


Figure 2.10: (A) Total escape scores for each site in the BA.1 spike and a zoomed-in plot showing the key escape sites. (B) Heatmap of mutation escape scores at key sites. Residues marked with X are the wild-type amino acids in BA.1. Amino acids not present in our libraries are shown in gray. An interactive version of this plot for the entirety of spike is at https://dms-

vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/NTD_5-7_escape_plot.html (C) Surface representation of spike coloured by the sum of escape scores at that site. Antibody 5-7 is shown in yellow in the inset. PDB ID: 7RW2. (D) Validation pseudovirus neutralization assays of the indicated BA.1 spike mutants against antibody 5-7. (E) Correlation between predicted IC50 values from deep mutational scanning (DMS) data versus the IC50 values measured in the validation assays in (F). Lower bound indicates that the antibody did not neutralize at the highest concentration tested in the validation neutralization assay. Site numbering in all plots is based on the Wuhan-Hu-1 sequence.

We next applied the full spike deep mutational scanning to S2 stem-helix targeting antibodies CC9.104 and CC67.105, which were isolated using both the SARS-CoV-2 and Middle East Respiratory Syndrome coronavirus (MERS-CoV) spike proteins as baits[103]. Both CC9.104 and CC67.105 broadly neutralize SARS-related coronaviruses, and CC9.104 also retains some potency against MERS-CoV. As expected, our deep mutational scanning showed that escape sites for both antibodies cluster in the S2 stem-helix region (Figure 2.11A-E; https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/CC67.105_escape_plot.html and https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/CC9.104_escape_plot.html). Our data also explain why only CC9.104 neutralizes MERS-CoV. The deep mutational scanning shows that the CC67.105 epitope centers on sites D1146, D1153 and F1156 (Figures 2.11B, 2.11D), and consistent with the deep mutational scanning, mutating these sites leads to complete escape in validation neutralization assays (Figure 2.11G). By contrast, the deep mutational scanning shows that while CC9.104's epitope also includes sites D1153 and F1156, mutations at site D1146 cause only modest or no escape (Figures 2.11C, 2.11E), and validation neutralization assays again confirm these deep mutational scanning results (Figure 2.11G). Notably, sites D1153 and F1156 are conserved between SARS-CoV-2 and MERS-CoV S2 stem-helix regions, but site D1146 is mutated to isoleucine in MERS-CoV (Figure 2.11F). Based on our deep mutational scanning data we expect changes at site D1146 to have significant effects on escape from the CC67.105 antibody but only modest effects on escape from the CC9.104 antibody. While our deep mutational scanning libraries do not include isoleucine at site

D1146 (which is found in MERS-CoV), we confirmed that other mutations at D1146 lead to complete escape from CC67.105 and do not substantially impact neutralization by CC9.104 (Figure 2.11G). Note that site D1163 is also mutated to isoleucine in MERS-CoV and both antibodies show some escape at that site, which may explain why CC9.104's potency against MERS-CoV is lower than against SARS-CoV-2.

The above deep mutational scanning of escape from the S2 antibodies emphasizes the difference between SARS-related coronavirus breadth and resistance to escape in SARS-CoV-2. Both CC9.104 and CC67.105 neutralize many diverse SARS-related coronaviruses, but Omicron sub-variants with mutations that lead to almost complete escape from these antibodies have been detected (e.g. D1153Y in BA.2.46 and BA.2.59). Therefore, even pan-sarbecovirus neutralizing antibodies can be escaped by mutational diversity within SARS-CoV-2, which emphasizes the importance of directly mapping escape mutations in SARS-CoV-2 in addition to assessing breadth across other natural SARS-related coronaviruses.

Figure 2.11 Antibody CC9.104 and CC67.105 escape mapping

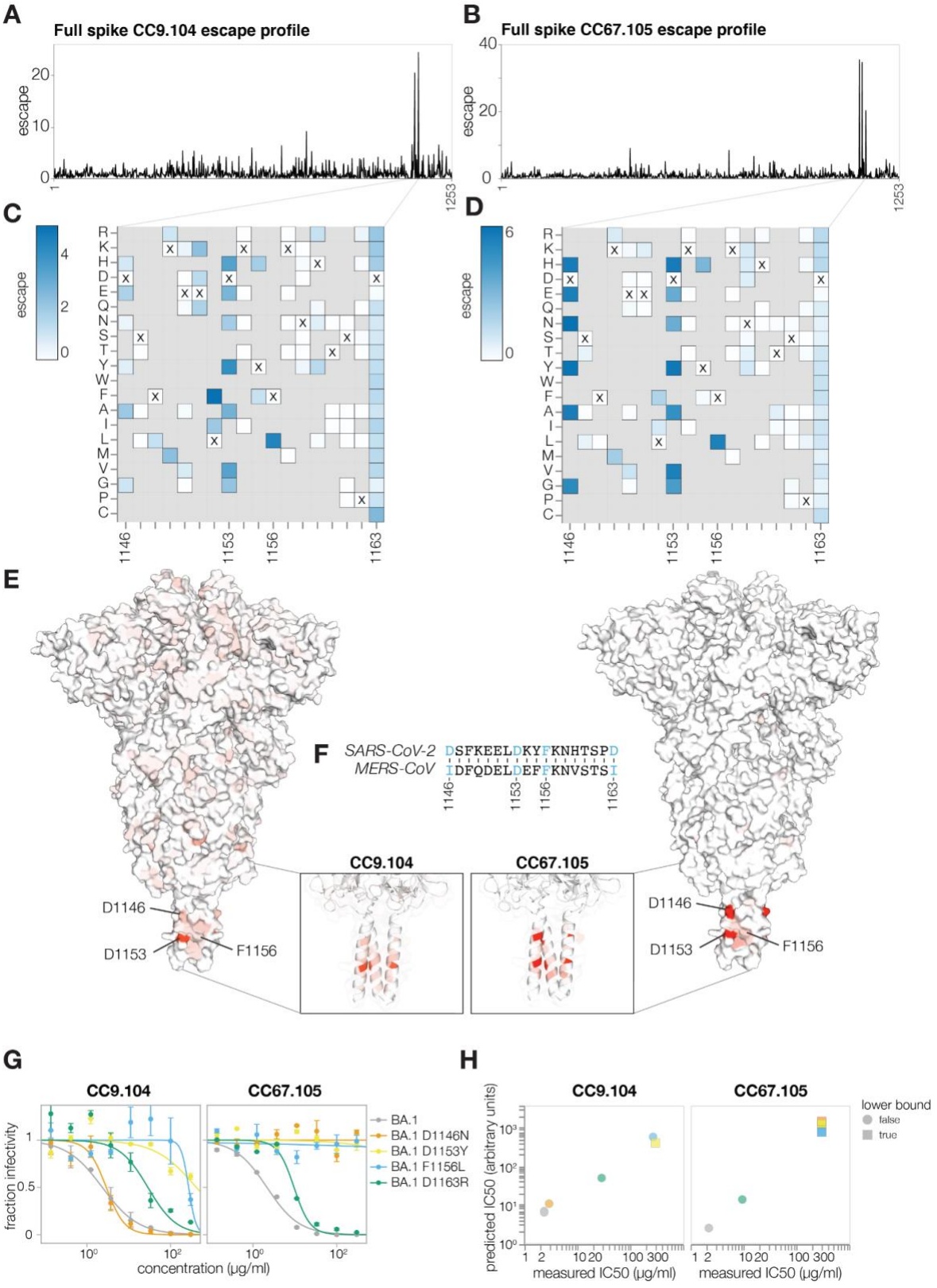


Figure 2.11: (A-B) Total escape scores for each site in the BA.1 spike for the CC9.104 (A) and CC67.105 (B) antibodies. (C-D) Escape heatmaps for the S2 stem-helix (sites 1146-1163) for CC9.104 (C) and CC67.105 (D) antibodies. Residues marked with X are the wild-type amino acids in BA.1 sequence. Amino acids that are not present in our libraries are shown in gray. Interactive heatmaps for the entirety of spike are at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/CC67.105_escape_plot.html and https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/CC9.104_escape_plot.html (E) Surface representation of spike coloured by the sum of escape scores at that site for CC9.104 (left) and CC67.105 (right) antibodies. Site 1163 is not resolved in the structure. PDB ID: 6XR8. (F) Alignment of SARS-CoV-2 and MERS-CoV spikes at sites 1146-1163. (G) Validation pseudovirus neutralization assay for CC9.104 (left) and CC67.105 (right) antibodies with BA.1 spike carrying the indicated mutations. (H) Correlation between predicted IC50 values from deep mutational scanning (DMS) data versus the IC50 values measured in the validation assays in (G). Lower bound indicates that the antibody did not neutralize at the highest concentration tested in the validation neutralization assay. Site numbering in all plots is based on the Wuhan-Hu-1 sequence.

To show that we can perform deep mutational scanning of the spikes from different SARS-CoV-2 strains, we mapped escape from the REGN10933 antibody using Delta spike deep mutational scanning libraries (Figure 2.12A-B). REGN10933 is a class 1 antibody that directly competes with ACE2 binding, and was part of REGN-COV2 therapeutic cocktail used early in the pandemic but has lost potency against Omicron variants[80,107,108]. Escape sites for REGN10933 mapped with our deep mutational scanning system overlapped with the antibody binding footprint and included previously described escape mutations[107,108] (Figure 2.12C) and correlated well with previously described yeast-based RBD deep mutational scanning data for this antibody (Figure 2.12C-D)[58].

Finally, to show that pseudovirus-based deep mutational scanning can be used to map escape from polyclonal sera, we mapped escape from two Delta breakthrough sera using our Delta libraries (Figure 2.13). Major escape sites for both sera resided in RBD class I and II epitopes (sites 417, 484 and 486) (Figure 2.13) as has been shown previously[85], and correlated well with changes in IC50 measured in conventional neutralization assays.

Figure 2.12 Antibody REGN10933 escape mapping using Delta deep mutational scanning libraries

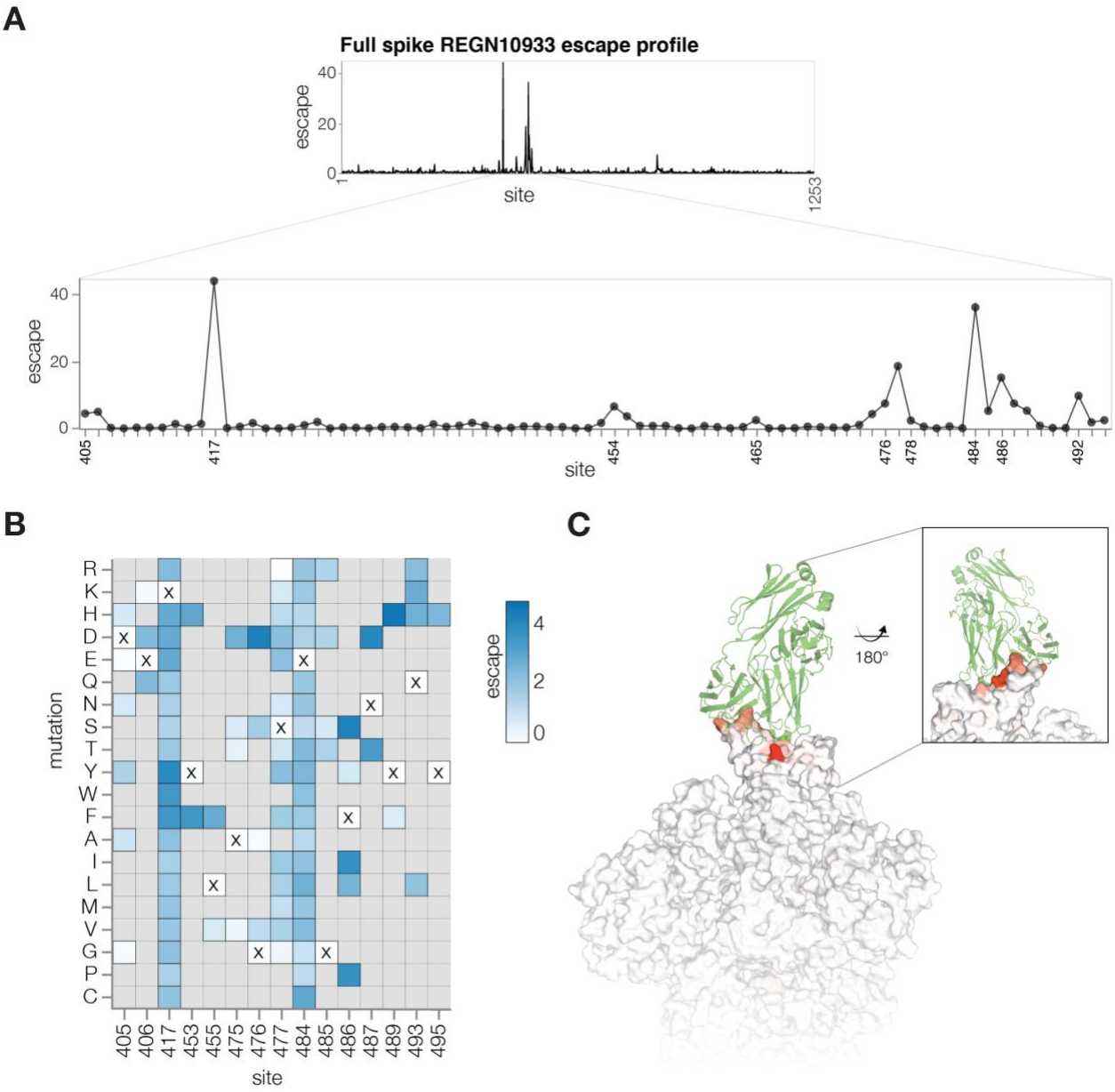


Figure 2.12: (A) Total escape scores for each site within Delta spike and a zoomed-in plot showing key escape sites. (B) Heatmap of mutation escape scores at key sites. Residues marked with X are the wild-type amino acids in Delta sequence. Amino acids not present in our libraries are shown in gray. An interactive version of this heatmap for the entirety of spike is at https://dms-vep.github.io/SARS-CoV-2_Delta_spike_DMS_REGN10933/REGN10933_escape_plot.html (C) Surface representation of spike coloured by sum of escape scores at that site. REGN10933 antibody is shown in green. PDB structures 6XDG and 6XM4 were aligned to make this figure. Site numbering in all plots is based on the Wuhan-Hu-1 sequence.

Figure 2.13 Delta breakthrough serum escape mapping using Delta deep mutational scanning libraries

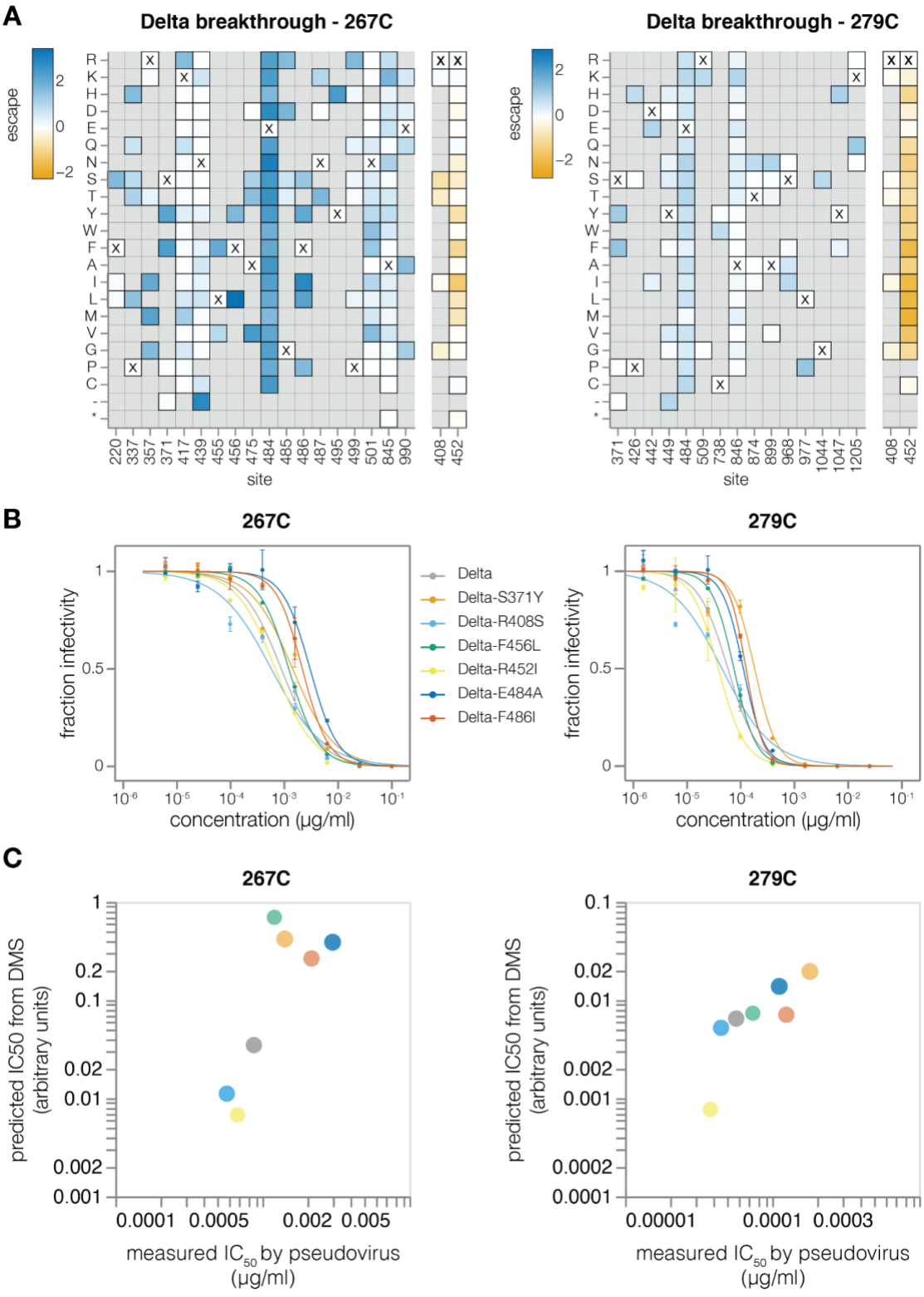


Figure 2.13: (A) Heatmap of mutation escape scores at key sites for Delta breakthrough sera 267C (left) and 279C (right). Sites shaded blue are escape mutations and sited shaded orange are sensitizing mutations. Residues marked with X are the wild-type amino acids in Delta sequence. Amino acids not present in our libraries are shown in gray. Interactive plots for these heatmaps can be found at https://dms-vep.github.io/SARS-CoV-2_Delta_spike_DMS_REGN10933/267C_escape_plot.html and https://dms-vep.github.io/SARS-CoV-2_Delta_spike_DMS_REGN10933/279C_escape_plot.html. (B) Validation pseudovirus neutralization assays of the indicated Delta spike mutants against the Delta breakthrough sera. (C) Correlation between predicted IC50 values from deep mutational scanning (DMS) data versus the IC50 values measured in the validation assays in panel (B). The points are colored as in panel (B). Site numbering in all plots is based on the Wuhan-Hu-1 sequence. Note that site 452 likely contains many sensitizing mutations because it is mutated in Delta relative to the original vaccine received by the individuals from which the sera is derived.

2.3.6 Functional effects of mutations on spike-mediated pseudovirus infection

Our deep mutational scanning also enables measurement of how mutations affect spike-mediated viral infection in the absence of antibodies. We can make these measurements by computing a functional score for each variant from its relative frequency in infectious spike-pseudotyped lentiviruses generated from our single-copy cell integrated cells versus VSV-G-pseudotyped lentivirus generated from the same cells (Figures 2.2B). Spike variants with negative functional scores are worse at mediating cellular infection than the parental unmutated spike, while variants with positive functional scores are better at mediating infection (Figure 2.5). To deconvolve the functional scores for the variants (which often contain multiple mutations) into the effects of individual mutations on spike-mediated entry, we used global epistasis models[76,77].

As expected stop-codon mutations to the BA.1 spike were highly deleterious for spike-mediated infection, whereas amino-acid mutations showed a wide range of effects ranging from slightly beneficial to roughly neutral to highly deleterious (Figure 2.14A; recall that our library design excludes many of the most deleterious amino-acid mutations). To test whether the mutations measured to have slightly beneficial effects actually improved spike-mediated infection, we chose five mutations that the deep mutational scanning indicated improved infection (Figure 2.14B), and generated pseudovirus mutants carrying these mutations. The validation experiments confirmed that all the tested mutations indeed slightly improved spike-

mediated infection (Figure 2.14C), validating that our deep mutational scanning can identify mutations that increase spike-mediated pseudovirus infection.

Figure 2.14 Functional effects of mutations on spike-mediated pseudovirus infection

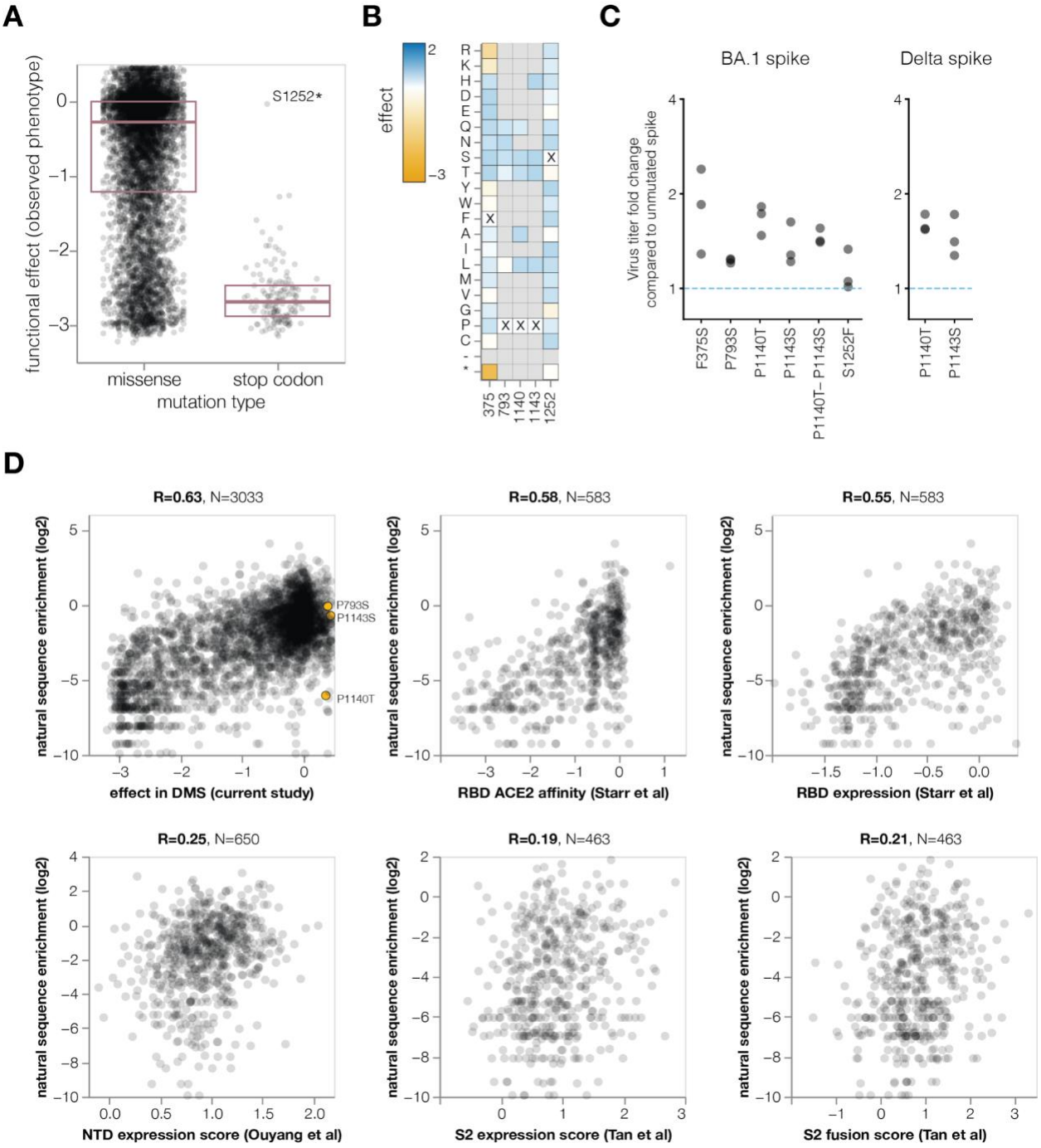


Figure 2.14: (A) Distribution of functional effects of mutations in BA.1 deep mutational scanning libraries. Negative values indicate mutations are deleterious for viral entry. The stop codon mutation with a neutral functional effect of ~0 is at the last codon of the spike used in our

experiments. (B) Heatmap showing functional effects at sites of mutations with beneficial functional effects that were chosen for validation assays in C. An interactive version of this heatmap for the entire spike is at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/muteffects_observed_heatmap.html (C) Fold change in virus entry titer for spike mutants relative to unmutated spike. There are three points for each mutant, reflecting triplicate measurements. (D) Correlation between enrichment of mutations during actual evolution of human SARS-CoV-2 and functional effects from our lentivirus-based deep mutational scanning or previous RBD expression or ACE2 affinity for yeast-based deep mutational scanning[88], and S2[63] or NTD[86] expression for mammalian display-based deep mutational scanning. Interactive plots that enable mouseovers and show correlations among experiments are at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/all_natural_enrichment_vs_dms.html. Interactive plots with correlations of enrichments during natural evolution calculated only from BA.1 sequences are at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/21K_natural_enrichment_vs_dms.html and have a lower correlation with the deep mutational scanning probably because a smaller dataset leads to less accurate estimates of enrichment during natural evolution.

To examine the relationship between the functional effects of spike mutations in the deep mutational scanning and actual evolution of human SARS-CoV-2, we determined the extent that mutations are enriched or depleted across a phylogenetic tree of all publicly available human SARS-CoV-2 sequences[97]. To do this, we calculated the number of independent observations of each mutation on the tree and compared these observed numbers to the expected numbers under neutrality as estimated from four-fold synonymous sites[109], analyzing only mutations expected to have ≥20 occurrences (see Methods for details). Our deep mutational scanning measurements of the effects of mutations on spike-mediated infection were reasonably correlated with the enrichment of mutations among actual sequences (Figure 2.14D), indicating our experiments at least partially reflect the functional selection actually shaping spike evolution. We performed a similar analysis for prior spike deep mutational scanning using yeast display of the RBD[88], or mammalian cell display of the NTD[86] or a region of S2[63] (Figure 2.14D). Our pseudovirus-based spike deep mutational scanning measurements

were more correlated with the enrichment of mutations during actual evolution than any of these prior cell-surface display deep mutational scanning studies, presumably because our experiments mimic the true biological function of spike better than cell-surface display experiments.

However, none of the mutations with positive deep mutational scanning functional scores that we validated to improve spike-mediated infection in a pseudovirus context are enriched during actual SARS-CoV-2 evolution (Figure 2.14D). We suggest that this is because there is some divergence between the selection pressure in our pseudovirus-based experiments and true natural selection on spike. For instance, mutations at sites P1140 and P1143, which are located at the beginning of the S2 stem-helix, could potentially destabilize the prefusion trimer leading to more rapid cell entry in a pseudovirus context but negatively affect spike stability in the context of actual human transmission. Nonetheless, our functional measurements still provide the most accurate large-scale measurements to date on the effects of mutations to spike, and should be useful for assessing which antibody-escape mutations are well enough tolerated to pose a plausible risk of emerging naturally. We also note that our experiments indicate that there are no further mutations to the BA.1 spike that improve pseudovirus titers to the same extent as the D614G mutation that fixed early in SARS-CoV-2's evolution in humans[110–112].

2.4 Discussion

We have developed a new deep mutational scanning system for assessing the antigenic and functional effects of mutations in the SARS-CoV-2 spike. This deep mutational scanning system is the first to measure how mutations to the entirety of spike affect cellular infection, and therefore enables the mapping of escape from antibodies targeting any part of the spike. Furthermore, our system directly quantifies how mutations affect antibody neutralization, and we show that these measurements correlate well with traditional pseudovirus neutralization assays.

We expect that the ability to directly measure neutralization as opposed to binding will be especially useful when applied to polyclonal sera, since the magnitude of how mutations affect neutralization versus binding can differ in a polyclonal context[83,113].

Using the new deep mutational scanning system, we have mapped mutations that escape monoclonal antibodies targeting the RBD, NTD, and S2 domains. Although we only characterized a few antibodies here, this work is a first step towards generating similar data for much larger antibody panels. For the RBD, such data has proven valuable for interpreting the antigenic phenotypes of the new SARS-CoV-2 variants constantly emerging in the human population[83,85]. The ability to generate comparable data for other spike domains should enable better antigenic assessment of new variants, which can help inform vaccine design and strain selection[114]. In addition, our work on the S2 antibodies emphasizes the distinction between breadth on natural viruses and potential for escape within SARS-CoV-2: an antibody that cross-neutralizes MERS-CoV and all tested SARS-related viruses is still escaped by mutations already present in some SARS-CoV-2 variants. Therefore, direct prospective mapping of escape mutations[58,84] should be useful for informing development of therapeutic antibodies targeting even regions of spike conserved across most natural viruses.

We also used the deep mutational scanning system to measure how mutations in spike affect its ability to mediate pseudovirus infection. These measurements complement existing deep mutational scanning datasets on how spike mutations affect the molecular phenotypes of ACE2 affinity[88], membrane fusion[63], and cell-surface expression[63,86,88]. None of these experimental measurements fully reflect how mutations affect actual viral fitness, which is an emergent property of many molecular phenotypes[115,116]. However, understanding how mutations affect molecular phenotypes is an important step towards interpreting evolutionary outcomes, and our pseudovirus-based system currently provides the deep mutational scanning measurements that best reflect natural selection on the spike of human SARS-CoV-2.

The deep mutational scanning system we describe here can also be straightforwardly extended to any virus with an entry protein amenable to lentiviral pseudotyping. This set of viruses includes other coronaviruses, influenza viruses, filoviruses, arenaviruses, and henipaviruses—all of which have receptor-binding and fusion proteins for which lentiviral pseudotyping provides a safe way to study cellular infection and antibody neutralization without requiring direct work with the actual pathogenic virus[117–121]. Deep mutational scanning of the entry proteins of all these viruses could provide valuable information for antigenic surveillance and vaccine design, since these proteins are the dominant target of neutralizing antibodies. However, we note that data generated by such a system could in principle be used to inform introduction of gain-of-function mutations into actual potential pandemic viral pathogens. We therefore suggest that advances in the high-throughput characterization of mutations to viral proteins should be coupled with thoughtful limits on any downstream experiments with actual replicating viruses[122] to ensure that safely generated information is used to benefit public health without creating new risks.

2.5 Limitations of this study

The full-spike lentiviral-based deep mutational scanning system described in this paper has both limitations and advantages compared to other systems to identify SARS-CoV-2 antibody escape mutations. Yeast-display based RBD deep mutational scanning[57,83,88] is probably more high-throughput (eg, Cao et al.[83] have applied it to thousands of antibodies), but is limited to only RBD antibodies and measures binding rather than neutralization. Passaging of spike-pseudotyped VSV in the presence of antibodies can effectively identify both individual escape mutations and combinations of mutations, and better simulates natural evolution[22,107], but does not provide comprehensive maps because the mutations that arise in any given experiment are stochastic. Passaging of live SARS-CoV-2 has similar benefits and limitations to the spike-pseudotyped VSV[123] but also involves biosafety concerns. The system we describe

here is comprehensive and applies to the full spike, but will only identify escape mutations introduced into the original library design.

An important caveat is that our system measures the effects of spike mutations on viral entry in a pseudotyped system in cell culture. While this system captures many of the actual constraints on spike, it does not capture all of them as evidenced by the fact that some of the mutations we identify that increase viral entry in our system do not appear to be favorably selected during the actual evolution of human SARS-CoV-2. Therefore, it is important to keep in mind that the measurements that are generated by our system come from a simplified experimental surrogate for the full role of spike during infection and transmission in humans.

2.6 Methods

Table 2.1 Primer sequences used for building deep mutational scanning libraries

Name	Sequence 5′-3′	Function
Mutagenesis primers		
BA.1 oPools	https://github.com/dms-vep/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/tree/main/library_design	Primers to add mutations to BA.1 spike
Delta oPools	https://github.com/dms-vep/SARS-CoV-2_Delta_spike_DMS_REGN10933/tree/main/library_design/results	Primers to add mutations to Delta spike
Spike amplification and joining PCR		
VEP_amp_for	CAGCCGAGCCACATCGCTC	For amplifying spike with flanking sequences from lentivirus backbone
3'rev_lib_LinJoin_KHDC	CGGAAGAGCGTCGTGTAGGGAAAG	
Spike barcoding primers		
ForInd_AddBC_2	gcggaactccactaggaacatttctctctcgaaTCTAGANNNNNNNNNNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAG	For adding barcodes to spike gene
5'for_lib_bcing	gcacgcgCAGCCGAGCCACATCGCTCA	
Neutralization standard barcoding primers		
5'for_lib_bcing	gcacgcgCAGCCGAGCCACATCGCTCA	For barcoding neutralization

Barcoding primer pool 1: VSVG_BC1 VSVG_BC2 VSVG_BC3 VSVG_BC4	gcggaactccactaggaacatttctctctcgaaTCTAGAtactttactactgcacAGATCGGAAGAGCGTCGTGTAGGGAAAGAG gcggaactccactaggaacatttctctctcgaaTCTAGAggaccattgcgacgtaAGATCGGAAGAGCGTCGTGTAGGGAAAGAG gcggaactccactaggaacatttctctctcgaaTCTAGAcctagccactagatggAGATCGGAAGAGCGTCGTGTAGGGAAAGAG gcggaactccactaggaacatttctctctcgaaTCTAGAatggagggagtctactAGATCGGAAGAGCGTCGTGTAGGGAAAGAG	standard
Barcoding primer pool 2 VSVG_BC5 VSVG_BC6 VSVG_BC7 VSVG_BC8	gcggaactccactaggaacatttctctctcgaaTCTAGAtagtgtaaacgccacgAGATCGGAAGAGCGTCGTGTAGGGAAAGAG gcggaactccactaggaacatttctctctcgaaTCTAGAccaacgcgtgaatcgcAGATCGGAAGAGCGTCGTGTAGGGAAAGAG gcggaactccactaggaacatttctctctcgaaTCTAGAatcgtatccatgggtaAGATCGGAAGAGCGTCGTGTAGGGAAAGAG gcggaactccactaggaacatttctctctcgaaTCTAGAggtcacgtgtctatatAGATCGGAAGAGCGTCGTGTAGGGAAAGAG	
Spike gene amplification for PacBio long-read sequencing		
PacBio_5pri_C_tag1	ctagccattcagagGCAGCCGAGCCACcTCGCTC	Used for amplifying spike and barcode for PacBio sequencing
PacBio_3pri_G_tag1	CGCTCAACCAGTACGAGCCGTAAGTTATGTAACGCGGAACTCCACgAGGAAC	
PacBio_5pri_G_tag2	ctagccattcagagGCAGCCGAGCCACgTCGCTC	
PacBio_3pri_C_tag2	CGCTCAACCAGTACGAGCCGTAAGTTATGTAACGCGGAACTCCACcAGGAAC	
PacBio_5pri_RND2	CTAGCCATTCAGAGGCAGCCGAG	
PacBio_3pri_RND2	CGCTCAACCAGTACGAGCCGTAAGTTATGTAAC	
Illumina barcode sequencing 1st round PCR primers		
IlluminaRnd1_For	CTCTTTCCCTACACGACGCTCTTCCGATCT	Round 1 primers for barcode sequencing
IlluminaRnd1_Rev3	CTGGAGTTCAGACGTGTGCTCTTCCGATCTgtccctattggcgttactatgggaacatacgtc	
Illumina barcode sequencing library round 2 PCR primers		
Rnd2ForUniversal	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	Universal illumina primer and indexing primer, where xxxxxxx indicates i7 index sequence
Indexing primer	CAAGCAGAAGACGGCATACGAGATxxxxxxxxGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	

2.6.1 Design of lentiviral backbone and spike gene nucleotide sequence optimization

The structure of lentiviral backbone is shown in Figure 2.2A. The plasmid map of the lentivirus backbone containing BA.1 spike is at https://github.com/dms-vep/SARS-CoV-

2_Omicron_BA.1_spike_DMS_mAbs/blob/main/library_design/reference_sequences/3282_pH2rU3_ForInd_Omicron_sinobiological_BA.1_B11529_Spiked21_T7_CMV_ZsGT2APurR.gb and the map for the Delta spike-containing backbone is at https://github.com/dms-vep/SARS-CoV-2_Delta_spike_DMS_REGN10933/blob/main/library_design/reference_sequences/pH2rU3_ForInd_sinobiological_617.2_Spiked21_CMV_ZsGT2APurR.gb. Note our BA.1 spike includes the EPE insertion after position 214 (Wuhan-Hu-1 numbering). The vector is based on a pHAGE2 lentiviral backbone in which we repaired the 3' LTR sequence[92], which allows us to re-rescue the pseudovirus from the cells in which lentiviral backbones have been integrated. The lentiviral backbone is non-replicative unless helper plasmids (Gag/Pol (NR-52517), Tat1b (NR-52518), and Rev1b (NR-52519)) are also transfected into the cells containing this backbone. Expression of the spike gene in the lentivirus backbone is driven both by inducible TRE3G promoter and by Tat1b. TRE3G promoter is activated by addition of doxycycline in the presence of the reverse tetracycline transactivator (rtTA), which is endogenously expressed in HEK-293T-rtTA cells. The spike gene has been codon optimized and lacks 21 amino acids in its cytoplasmic tail. The cytoplasmic tail deletion has been previously shown to significantly increase pseudovirus titers[124,125]. For spike sequence codon optimization we tested a large panel of optimized sequences and found that virus titers can vary between codon optimizations by as much as 100-fold. While we did observe some titer differences when virus was generated from just transfection reactions using lentiviral backbones and helper plasmids, these differences were even more noticeable when virus was generated from cells with integrated spike-carrying lentiviral backbones. Of the tested codon optimizations we found that the sequence optimized spike from SinoBiological (VG40609-UT) gave by far the best virus titers; we therefore based all variant sequences on the original SinoBiological optimization. In addition to the inducible promoter and spike gene, the backbone also has a CMV promoter that drives expression of the ZsGreen gene linked by a T2A linker to the puromycin resistance gene. ZsGreen is used as a

reporter gene to detect pseudovirus infection and the puromycin resistance gene is used as a selection marker for cells with successfully integrated lentiviral backbones.

2.6.2 Design of spike mutations to include in BA.1 and Delta full spike deep mutational scanning libraries

We aimed to create a library with mutations that would result in mostly functional spike proteins and include the mutations likely to be important for the antigenic evolution of spike. To this end, for BA.1 and Delta deep mutational scanning libraries we included the following types mutations: (1) mutations (nonsynonymous changes and deletions) observed in spike sequences deposited on GISAID database, (2) mutations that reoccur in spike phylogeny independently multiple times, (3) all possible amino acid changes at sites in spike that show positive selection. Specifically for the BA.1 library, we also included all possible amino acid changes for sites that are mutated in the BA.1 spike relative to Wuhan-Hu-1.

The following criteria were used to select the above described mutations for BA.1 library: nonsynonymous mutations need to be present in GISAID database >16 times, deletions need to occur in the NTD and be observed on GISAID database >300 times, nonsynonymous mutations need to reoccur on spike phylogenetic tree independently at least 21 times. To get all spike mutations observed in GISAID deposited sequences we used a CoVsurver curated spike amino acid frequency table (with sequences deposited up to January-31-2022)[96]. To get independently recurring spike mutation counts we used pre-built SARS-CoV-2 phylogenies from UShER[97]. Information on sites in spike undergoing positive selection was taken from taken from table here https://raw.githubusercontent.com/spond/SARS-CoV-2-variation/master/windowed-sites-fel-2021-07.csv which was built using methods described in Maher et al.[98]. The full list of mutations included in the BA.1 library can be found at

https://github.com/dms-vep/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/blob/main/library_design/results/aggregated_mutations.csv .

The following criteria were used to select the above described mutations for the Delta library: nonsynonymous mutations and deletions need to be observed on GISAID database more than once and nonsynonymous mutations need to reoccur on spike phylogenetic tree independently more than 7 times. To get all spike mutations observed in GISAID deposited sequences we aligned all spike sequences deposited on GISAID up to July-26-2021 and extracted mutation frequency counts. Independently recurring spike mutations and positively selected sites were identified as described for BA.1 library above. The full list of mutations included in the Delta library can be found at https://github.com/dms-vep/SARS-CoV-2_Delta_spike_DMS_REGN10933/blob/main/library_design/results/aggregated_mutations.csv .

2.6.3 Design of primers for BA.1 and Delta spike mutagenesis

For each set of mutations described in the section above we designed separate primer pools: (1) a pool of primers for observed mutations, (2) a pool of primers for recurrent mutations, (3) a pool of primers for positive selection site mutations, and (4) a pool of primers that would cover changes at multiple positive selection sites if those positive selection sites are close enough to each other so that the primers in the pool (3) would overlap. For the BA.1 library we also designed primers that would introduce multiple amino acid deletions at recurrent deletion regions described in[126] and included them in the observed mutation primer pool. Also for the BA.1 library, the set of primers that cover all possible amino acid changes at the sites already mutated in BA.1 was pooled with the positive selection site primer pool.

All primer pools were ordered from Integrated DNA Technologies as oPools. Scripts for designing the BA.1 library primer pools and the resulting oPools that were ordered are at

https://github.com/dms-vep/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/tree/main/library_design and scripts for designing the Delta library primer pools are at https://github.com/dms-vep/SARS-CoV-2_Delta_spike_DMS_REGN10933/tree/main/library_design .

2.6.4 Design of full-spike deep mutational scanning plasmid libraries

Making of the plasmid libraries for deep mutational scanning required the following three steps (1) mutagenesis of the spike gene, (2) barcoding of the mutagenised spike sequence, and (3) cloning of the mutagenised and barcoded spike into the lentiviral backbone-carrying plasmid.

Spike mutagenesis was carried out by first amplifying BA.1 or Delta spike gene sequence from a plasmid carrying lentiviral backbone with a codon optimized spike sequence (see section 'Design of lentiviral backbone and spike gene nucleotide sequence optimization' for plasmid maps). Spike sequence was amplified using 'Spike amplification' primers from Supplementary Table 1 with the following PCR conditions: 1.5 µl of 10µM forward primer, 1.5 µl of 10 µM reverse primer, 10 ng of amplified spike gene template, 25 µl of KOD polymerase (KOD Hot Start Master Mix, Sigma-Aldrich, Cat. No. 71842), and water for the final volume of 50 µl. PCR cycling conditions were as follows:

1. 95ºC for 2 min
2. 95ºC for 20 s
3. 62ºC for 15 s
4. 70ºC for 2 min (return to step 2 for another 19x cycles)
5. Hold at 4ºC

Amplified spike sequence was first gel-purified using NucleoSpin Gel and PCR Clean-up kit (Takara, Cat. No. 740609.5) and then further purified using Ampure XP beads (Beckman Coulter, Cat. No. A63881) at 1:2.6 sample to bead ratio.

Next the purified spike template was used in mutagenesis PCR using protocol described previously in Bloom, 2014[45] with a few modifications (see also https://github.com/jbloomlab/CodonTilingPrimers for general background on this protocol). Forward and reverse primers for mutagenesis PCR were pooled into separate pools at 1 : 2 : 2 : 0.2 per primer molar ratio between observed primer pool : recurrent primer pool : positive selection primer pool : paired positive selection primer pool. The pooling ratios are determined by the fact that recurrent and positively selected sites may be more antigenically and structurally important for spike. Two independent mutagenesis reactions were performed for each spike creating two independent biological library replicates (which means that they will have a unique set of barcodes and a unique set of mutation combinations in spike). For BA.1 libraries we performed two rounds of mutagenesis with the first round consisting of 8 mutagenic PCR cycles followed by the second round of 10 mutagenic PCR cycles. For Delta libraries one biological replicate consisted of a single round of 10 mutagenic PCR cycles and the second biological replicate consisted one round of 8 and another round of 10 mutagenic PCR cycles. Each mutagenesis PCR reaction was divided into forward and reverse reactions based on whether the reaction used the forward or reverse pool of mutagenesis primers. The PCR conditions for mutagenesis were as follows: 1.5 µl of 5 µM of forward or reverse mutagenesis primer pool, 1.5 µl of 5 µM of forward or reverse 'Spike amplification and joining PCR' primer (see supplementary table 1), 1.5 µl of 3 ng/µl linearised spike template, 4 µl of water, and 15 µl of KOD. The PCR cycling conditions were as follows:

1. 95ºC, 2 min
2. 95ºC, 20 s
3. 70ºC, 1 s
4. 54ºC, 20 s, cooling at 0.5ºC/s
5. 70ºC, 100 s (return to step 2 for the number of cycles described above)
6. 4ºC, hold

Between each mutagenic PCR round 20 cycles of joining PCR were performed. For joining PCR we used 4 µl each of 1:4 diluted forward and reverse mutagenesis reactions, 1.5 µL each of forward and reverse joining primers (see 'Spike amplification and joining PCR' primers in supplementary table 1), 4 µl of water, and 15 µl of KOD. PCR cycling conditions were identical to the mutagenesis PCR conditions. Joined PCR mutagenesis products were gel and Ampure XP purified after each joining reaction.

After the spike sequence was mutagenised we performed a barcoding PCR that appended a random 16 nucleotide barcode sequence downstream of the spike gene stop codon. We chose 16 nucleotide barcodes as it allows for a total of 416 unique barcoded variants, which is much greater diversity of barcodes than the final size of our deep mutational scanning plasmid libraries and therefore limits potential barcodes duplications. For barcoding 'Spike barcoding' primers from supplementary table 1 were used with the following PCR conditions 1.5 µl of 10µM forward primer, 1.5 µl of 10 µM reverse primer, 30 ng of the mutagenised spike gene template, 25 µl of KOD polymerase, and water for final volume of 50 µl. PCR cycling conditions were as follows:

1. 95ºC, 2 min
2. 95ºC, 20 s
3. 70ºC, 1 s
4. 55.5ºC, 20 s, cooling at 0.5ºC/s
5. 70ºC, 2 min (return to step 2 for another 9x cycles)
6. 4ºC hold

The mutagenised and barcoded spike was then cloned into lentiviral backbone-containing plasmid. First, we digested a lentiviral backbone containing plasmid using MluI and XbaI restriction sites. The map of plasmid used for vector digestion is at https://github.com/dms-vep/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/blob/main/library_design/reference_sequences/other_plas

mid_maps_for_library_design/3137_pH2rU3_ForInd_mCherry_CMV_ZsGT2APurR.gb . Digested vector was gel and Ampure XP purified. We then used 1:3 insert to vector ratio in a 1 hour Hifi assembly reaction using NEBuilder HiFi DNA Assembly kit (NEB, Cat. No. E2621). After the HiFi assembly, we Ampure XP purified the reaction and eluted it in 20 µl of water (note that elution in water as opposed to elution buffer enhances the subsequent electroporation efficiency). We used 1 µl of the purified HiFi product to transform 20 µl of 10-beta electrocompetent E. coli cells (NEB, C3020K). We performed 10 electroporation reactions to get a final count of > 2 million CFUs per library and plated transformed cells out on LB+ampicillin plates. We aim to make plasmid library from a much greater number of CFUs than the number of variants in our final virus libraries to minimize barcode duplication, as explained in the next section. About 16 hours after transformation bacterial colonies were scraped using liquid LB+ampicillin and plasmid stocks were prepared using QIAGEN HiSpeed Plasmid Maxi Kit (Cat. No. 12662). The final structure of the lentiviral genome with mutagenised spike cloned into it is shown in Figure 2.2A.

2.6.5 Production of cell-stored spike deep mutational scanning libraries

Production of cell-stored deep mutational scanning libraries required the following steps: (1) production of VSV-G-pseudotyped lentivirus, (2) infection of rtTA-expressing cells with VSV-G-pseudotyped virus, and (3) selection for transduced cells. These steps are illustrated in Figure 2.2B.

To generate VSV-G-pseudotyped virus for each library we plated 0.5 million HEK-293T cells per well in eight wells of two 6-well tissue culture dishes. Note we aim to produce VSV-G-pseudotyped virus stocks that have a greater number of infectious particles than the number of colonies scraped for plasmid libraries in order to not introduce any bottleneck on barcodes at this stage. The next day we transfected 0.25 µg of each helper plasmid (Gag/Pol, Tat1b, and Rev1b), 0.25 µg of VSV-G expression plasmid (https://github.com/jbloomlab/SARS-CoV-2-

BA.1_Spike_DMS_validations/blob/main/plasmid_maps/29_HDM_VSV_G.gb) and 1 µg of mutagenised and barcoded spike containing lentiviral vector (described in the section above). Transfections were done using BioT reagent (Bioland Scientific, Cat. No. B01-02) according to the manufacturer's instructions. 48 hours post transfection supernatants from each well were pooled, filtered through a surfactant-free cellulose acetate 0.45 µm syringe filter (Corning, Cat. No. 431220), and stored at -80ºC. VSV-G-pseudotyped viruses were titrated as described in Crawford et al.[70].

Next we infected HEK-293T-rtTA cells with the generated VSV-G-pseudotyped virus. The number of infectious virus units used in these infections allowed us to bottleneck the library size at the desired final variant number. For BA.1 libraries we attempted to bottleneck libraries at 100,000 variants and for Delta libraries we bottlenecked the libraries at 50,000 variants. Notably, we used a substantially lower number of variants to infect cells compared to the possible diversity of variants in our plasmid libraries. This allows us to limit any potential duplication of barcodes between different variants due to recombination in the lentivirus genome[71–73]. Note for BA.1, libraries Lib-1 and Lib-2 originate from the same mutagenised lentiviral backbone plasmid stock but independent VSV-G virus infections and Lib-3 originates from independent mutagenised plasmid library stock. For Delta libraries Lib-1 and Lib-2 are both from independent mutagenised spike plasmid stocks. Infections were performed at MOI < 0.01 (in order to ensure that only a single spike variant is integrated in each cell), which was verified 48 hours after infection using fluorescence-activated cell sorting by detecting ZsGreen expression from the lentiviral backbone. After MOI was verified, we expanded cells for another 48 hours and then started puromycin selections to select for cells with successfully integrated lentivirus genomes. Selection was done using 0.75 µg/ml of puromycin with a fresh change of puromycin-containing D10 (see ‘Cell lines’ section below) every 48 hours. Selections were terminated when visual inspection using a fluorescent microscope indicated that all cells express ZsGreen (approximately 6-8 days). After puromycin selection was finished we

expanded cells for another 48 hours in fresh D10 and froze cell aliquots in tetracycline-free FBS (Gemini Bio, Cat. No. 100-800) containing 10% DMSO. Frozen cell aliquots were stored in liquid nitrogen long-term.

2.6.6 Generation of spike and VSV-G-pseudotyped viruses from cell-stored spike deep mutational scanning libraries

To generate spike-pseudotyped viruses from cell-stored deep mutational scanning libraries we plated 100 million library-containing cells per 5-layer flask (Corning Falcon 875cm² Rectangular Straight Neck Cell Culture Multi-Flask, Cat. No. 353144) in 150 ml of D10 without phenol red supplemented with 1 µg/ml for doxycycline (which allows to induce spike expression ahead of pseudovirus production). 24 hours after plating we transfected cells with 50 µg of each helper plasmid (Gag/Pol, Tat1b, Rev1b) using BioT reagent according to manufacturer's instructions. 48 hours post transfection cell supernatant was collected and filtered through a 0.45 µm SFCA Nalgene 500mL Rapid-Flow filter unit (Cat. No. 09-740-44B). Filtered supernatant was then concentrated by spinning at 4ºC 3000 rcf for 30 min using Pierce Protein Concentrator (ThermoFisher, 88537). Virus aliquots were stored long-term at -80ºC. Titers for spike-pseudotyped libraries titrated on HEK-293T-ACE2 cells ranged between 0.5x106 - 2x106 transcription units per ml after concentrating the virus by ~15-20 fold. We note that the library virus needs to be concentrated because the library virus titers are significantly lower than the wild-type single integrant titers (shown in Figure 2.1) due to the presence of deleterious mutations in many of the library variants.

To generate VSV-G-pseudotyped viruses (for functional selection and long-read PacBio sequencing) from cell-stored deep mutational scanning libraries we plated 60 million library-containing cells per 3-layer flask (Corning Falcon 525cm² Rectangular Straight Neck Cell Culture Multi-Flask, 353143) in 90 ml of D10 without phenol red (note we do not add doxycycline in this case). 24 hours after plating we transfected cells with 30 µg of each of the

helper plasmid (Gag/Pol, Tat1b, Rev1b) and 18.75 µg of VSV-G expression plasmid using BioT reagent according to manufacturer's instructions. 32-36 hours post transfection cell culture supernatant was collected and filtered through a 0.45 µm SFCA Nalgene filter unit. Filtered supernatant was then concentrated by spinning at 4ºC 3000 rcf for 30 min using Pierce Protein Concentrator. Virus aliquots were stored long-term at -80ºC. Titers for VSV-G-pseudotyped libraries titrated on HEK-293T-ACE2 cells ranged between 10x106 - 30x106 transcription units per ml after concentrating the virus by ~15 fold.

2.6.7 Long-read PacBio sequencing of barcoded spike variants in deep mutational scanning libraries

Long-read PacBio sequencing was used to acquire reads spanning the spike and the random 16 nucleotide barcode sequences. To prepare amplicons for PacBio sequencing we infected 1 million HEK-293T cells with ~30 million VSV-G-pseudotyped lentiviruses carrying the deep mutational scanning libraries. This number of viruses is significantly greater than the expected number of variants in the library, which allows us to achieve high variant coverage, avoid bottleneck of barcode diversity and correct for any potential PCR or sequencing errors. 12-15 hours after infection cells were trypsinized, washed with PBS and non-integrated lentiviral genomes were recovered using QIAprep Spin Miniprep Kit (Cat. No. 27106X4)[47,127]. We use non-integrated viral genomes as our sequencing templates because they are the more abundant forms of the lentiviral genome than the integrated proviruses[128–131]. Elution volume for the miniprep was adjusted to 144 µl . Next we performed two rounds of PCR to amplify the region in the lentivirus genome spanning the spike and the random 16 nucleotide barcode. In the first round of PCR we use primers containing single nucleotide tags, which allow us to later detect strand exchange that may occur during PCR amplification. To limit strand exchange during PCR (which would disrupt barcode/spike variant linkage) we also minimize the number of PCR cycles performed and do multiple PCR reactions per sample[132,133]. Each sample was split

into eight PCR reactions, four of which use ‘tag_1’ forward and reverse primers and four of which use ‘tag_2’ forward and reverse primers from the ‘Spike gene amplification for PacBio long-read sequencing’ primer set in Supplemental Table 1. PCR reaction conditions were as follows: 1 µl of forward primer, 1 µl of reverse primer, 20 µl of KOD, and 18 µl of sample. PCR cycling conditions for round 1 PCR were as follows:

1. 95ºC for 2 min
2. 95ºC for 20 s
3. 70ºC for 1 s
4. 60ºC for 10 s (ramp 0.5ºC/s)
5. 70ºC for 2.5 min (go to 2 for another 7 cycles)
6. 70ºC for 5 min
7. 4ºC hold

After the first PCR round we pooled all reactions for each sample and purified them using Ampure XP beads with 1:0.8 beads to sample ratio and the PCR product was eluted in 84 µl of elution buffer. Eluted PCR product was divided into four PCR tubes and the second round of PCR was performed using ‘RND2’ forward and reverse primers from the ‘Spike gene amplification for PacBio long-read sequencing’ primer set in the Supplemental Table 1. PCR reaction conditions were as follows: 2 µl of forward primer, 2 µl of reverse primer, 25 µl of KOD, and 21 µl of purified sample. PCR cycling conditions were the same as for the round 1 PCR for a total of 10 PCR cycles. PCR reactions for each sample were pooled, purified using Ampure XP beads with 1:0.8 beads to sample ratio, and eluted in 27 µl of elution buffer. Barcodes were attached to each sample using sample SMRTbell prep kit 3.0 before multiplexing. Multiplexed SMRTbell libraries were then bound to polymerase using Sequel II Binding Kit 3.2 and sequenced with PacBio Sequel IIe sequencer with a 20-hour movie collection time.

2.6.8 Antibody and sera escape mapping using full spike deep mutational scanning libraries

For antibody and sera escape mapping we used between 4-15 times more infectious virions than the estimated total number of barcodes in a deep mutational scanning library. Using significantly more infectious virions relative to the number of variants per library avoids bottlenecking by having multiple copies of each variant. Note that we expect there to be several fold more lentiviral genomes per selection experiment than the amount of infectious units used because we are recovering the non-integrated viral genomes for sequencing, which are more abundant than integrated proviral DNA[128–131] on which our library virus titers are based. For each antibody escape mapping experiment we made a master mix of library spike-pseudotyped virus mixed with VSV-G pseudotyped neutralization standard (described below). Neutralization standard was added at 1-2% of the total virus titer used in the experiment. Virus master mix was then aliquoted into eppendorf tubes to which either different mounts of antibody or no antibody was added. For LY-CoV1404, CC9.104 and CC67.105 antibodies we performed selection experiments at 3 concentrations, starting with IC99 concentration predetermined using standard pseudovirus neutralization assay and then increasing this concentration 4 fold and 16 fold. We start with IC99 concentration intending that around 1% of the library will be able to escape antibody selection. We use additional concentrations as it helps us to cover a greater dynamic concentration range in cases where the exact IC99 value is difficult to determine. Also, the use of multiple concentrations enables more precise mutation-escape predictions by the biophysical model used to decompose single-mutation effects[78]. For LY-CoV1404 starting concentration was 0.654 µg/ml, for CC9.104 - 68 µg/ml, for CC67.105 - 52.5 µg/ml. For the REGN10933 we started at IC99.5 at 0.146 µg/ml and also increased that concentration by 4 fold and 16 fold. For the NTD 5-7 antibody, which does not fully neutralize the virus, we started with >IC96 concentration at 150 µg/ml and then increased that concentration by 2 fold. For Delta

breakthrough sera we started at 0.1164 and 0.00352 sera dilutions for 267C and 279C samples, respectively. For 267C we increased serum concentration by 2.5 fold and for 279C we increased serum concentration by 4 and 8 fold. Virus was mixed with the antibody or serum by inverting tubes several times, spun down at 300 g, and incubated at 37ºC for 1 h. After incubation virus and antibody mix or no antibody control were used to infect approximately 0.5 million target cells, which were plated a day before in D10 supplemented with 2.5 µg/ml of amphotericin B (Sigma, Cat. No. A2942) (which increases viral titers as shown in Figure 2.1C). The target cell line for different antibodies is determined by whether an antibody is able to neutralize pseudovirus on that cell line. We have previously described this phenomena in Farrell et al.[134] where we show that non-ACE2 competing antibodies do not fully neutralize pseudovirus on ACE2 overexpressing cells. While testing antibodies for the current study we also noticed that some S2-targeting antibodies are also not affected by ACE2 overexpression. Therefore, for LY-CoV1404, CC9.104 and CC67.105 antibodies we used HEK-293T-ACE2 as target cells but for NTD-targeting 5-7 antibody we used HEK-293T-ACE2-medium cells. For REGN10933 and Delta breakthrough sera we used HEK-293T-ACE2-TMPRSS2 as target cells, because TMPRSS2 overexpression increases Delta pseudovirus titers. 12-15 hours after infection cells were trypsinized, washed with PBS and non-integrated lentiviral genomes were recovered using QIAprep Spin Miniprep Kit and eluted in 21 µl of Qiagen elution buffer. Barcode reads for each sample were then prepared for Illumina sequencing using a method described in ‘Barcode amplicon preparation for Illumina sequencing’ section below.

2.6.9 Functional selections using full spike deep mutational scanning libraries

To perform functional spike selections we infected 1 million HEK-293T-ACE2 cells with 1-2 millions of the spike or VSV-G-pseudotyped viruses produced from deep mutational scanning library carrying cells (described earlier). As for antibody selections, the amount of virus used is greater than the number of variants in each library which limits potential bottlenecking of

the library barcodes. 12-15 hours after infection cells were trypsinized, washed with PBS and non-integrated lentiviral genomes were recovered using QIAprep Spin Miniprep Kit. Barcode reads for each sample were then prepared for Illumina sequencing using methods described in 'Barcode amplicon preparation for Illumina sequencing' section below.

2.6.10 Barcode amplicon preparation for Illumina sequencing

To prepare barcode reads for Illumina sequencing we performed two rounds of PCR. In the first round of PCR we used primers that align to Illumina Truseq Read 1 primer site located directly upstream of the barcode in the lentiviral backbone and a primer annealing downstream of the barcode containing an overhand with Illumina Truseq Read 2 sequence (see 'Illumina barcode sequencing 1st round PCR primers' in the Supplemental Table 1). Conditions for the first round PCR were as follows: 1 µl of 10uM forward primer, 1 µl of 10uM reverse primer, 26 µl of KOD, and 24 µl of miniprepped sample DNA. PCR cycling conditions for round 1 PCR were as follows:

1. 95ºC for 2 min
2. 95ºC for 20 s
3. 70ºC fro 1 s
4. 58ºC for 10 s, cooling at 0.5ºC per s
5. 70ºC 20 s (return to step 2 for another 27 cycles)
6. 4ºC hold

PCR reactions were purified with Ampure XP beads using 1:3 sample to beads ratio and eluted in 37 µl of Qiagen elution buffer. Second round of PCR used primers primer annealing to the Illumina Truseq Read 1 primer site with P5 Illumina adapter overhang and reverse primers from the PerkinElmer NextFlex DNA Barcode adaptor set, which anneal to Truseq Read 2 site and contain P7 Illumina adapter and i7 sample index. Conditions for the second round PCR were as follows: 1.5 µl of 10uM universal primer, 1.5 µl of 10uM indexing primer, 25 µl of KOD, and 20

ng of first round PCR product. PCR cycling conditions were the same as the first round PCR for a total of 20 cycles. After the second PCR round all samples were pooled at desired ratios and gel and Ampure XP bead purified. Barcode amplicons were sequenced using NextSeq 2000 with either P2 or P3 reagent kits.

2.6.11 Production of barcoded neutralization standard

To make the neutralization standard that we add into our deep mutational scanning libraries we used the same general barcodinging approach as described above for the deep mutational scanning plasmid library generation with a few important differences. The lentiviral backbone used for the neutralization standard consists of TRE3G inducible mCherry protein and CMV promoter driven ZsGreen. The plasmid map of the template backbone is at https://github.com/dms-vep/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/blob/main/library_design/reference_sequences/other_plasmid_maps_for_library_design/2871_pH2rU3_ForInd_mCherry_CMV_ZsG_NoBC_cloningvector.gb . Note this backbone does not encode any viral glycoproteins and to rescue VSV-G-pseudotyped virus we provide VSV-G expression plasmid in trans. We amplified mCherry plasmid from the lentiviral template and barcoded it in two independent PCR reactions using 2 sets of primers containing 4 unique barcodes (see 'Neutralization standard barcoding primers' in supplementary table 1). Importantly, the unique barcodes were balanced so that there is a unique nucleotide at each position of the 16-nucleotide barcode between each of the four barcoding primers in a PCR reaction. Furthermore, the 8 barcoding primers are unique to the neutralization standard and are not present in any of our deep mutational scanning libraries. The PCR for barcoding was done the same way as described for deep mutational scanning plasmid library production and both PCR reactions were pooled together before HIfi assembly into the lentiviral backbone. VSV-G-pseudotyped lentiviral particles were then produced as described above (Production of cell-stored spike deep mutational scanning

libraries) for the spike libraries, except using this mCherry-containing barcoded backbone as the lentiviral backbone. This VSV-G-pseudotyped virus was then used to infect HEK-293T-rtTA cells at low MOI. Successfully transduced HEK-293T-rtTA were then selected by flow activated fluorescence sorting based on ZsGreen expression and expanded. We then generated VSV-G-pseudotyped neutralization standard by transfecting helper plasmids and VSV-G expression plasmid in the same way as described for deep mutational scanning library virus rescues. Note, that we use the neutralization standard generated from the integrated cells as opposed to the original transfection in order to prevent lentiviral backbone-containing plasmid contamination of the virus stocks that can occur when viruses are produced from transfections.

2.6.12 Validation of deep mutational scanning by pseudovirus titration and neutralization

Spike genes carrying desired mutations were cloned by performing PCR reactions with partially overlapping desired mutation-containing primers followed by HiFi assembly. HDM_omicron_B11529_IDTDNA plasmid was used as the template for PCR. The map of the plasmid can be found at https://github.com/jbloomlab/SARS-CoV-2-BA.1_Spike_DMS_validations/blob/main/plasmid_maps/3277_HDM_omicron_B11529_IDTDNA.gb. All plasmid sequences were verified using full plasmid sequencing by Primordium. Mutated spike plasmids or VSV-G expression plasmid were then used to generate and titrate pseudoviruses as described in Crawford et al.[70] except that the backbone used for virus generation was pHAGE6_Luciferase_IRES_ZsGreen and which also only required Gag/Pol helper plasmid for virus rescues. Note, for the spike variants cloned to validate functional selections we performed three replicate virus rescues for each variant and each rescue was done using an independent plasmid preparation for that spike variant.

BA.1 spike variants rescued for functional selection validation were titrated on HEK-293T-ACE2 and Delta spike variants were titrated on HEK-293T-ACE2-TMPRSS2 cells. We

performed duplicate serial dilutions using supernatants collected from the virus rescues and measured luciferase expression at each dilution using Bright-Glo Luciferase Assay System (Promega, E2610). Virus titers were calculated as relative light units (RLU) per µl for each dilution and taking the average RLU/µl values across dilutions within a linear range. For spike variants used to validate antibody escape experiments virus titration was performed in the same way using the same target cells as the neutralization assays were performed in (see below).

For pseudovirus neutralization 12.5 thousand target cells were plated into poly-L-lysine coated, black-walled, 96-well plates (Greiner 655930) in D10 supplemented with 2.5 µg/ml of amphotericin B. For neutralization assays using LY-CoV1404, CC9.104 or CC67.105 antibodies we used HEK-293T-ACE2 as target cells, for REGN10933 and Delta breakthrough sera we used HEK-293T-ACE2-TMPRSS2 as target cells, and for NTD 5-7 mAb we used HEK-293T-ACE2-medium as target cells. The use of different cell lines for each antibody is determined by the ability of that antibody to neutralize the virus on that cell line as described previously. Next day we prepared replicate serial dilutions for each antibody. Starting concentration for each antibody was as follows: LY-CoV1404 – 4 µg/ml, CC9.104 and CC67.105 – 300 µg/ml , 5-7 – 96 µg/ml, REGN10933 – 6 µg/ml. Starting dilutions for Delta breakthrough sera were 5 fold dilution for 267C sera and 12.5 fold dilution for 279C sera. Serial dilutions were then mixed with pseudovirus and incubated for 1 h at 37ºC. After incubation the virus-antibody or virus-sera mix was transferred onto the target cells. 48-55 h after infection Bright-Glo Luciferase Assay System (Promega, E2610) was used to measure luciferase activity. Fraction infectivity for each antibody dilution was calculated by subtracting background readings and dividing RLU values in the presence of antibody by RLU values in the absence of it. Neutralization curves were then plotted by fitting a Hill curve to fraction infectivity values using neutcurve software (https://jbloomlab.github.io/neutcurve/, version 0.5.7). Neutcurve package was also used to extract target ICx values from the fitted neutralization curves.

Code for plotting virus titers and neutralization curves from this paper can be found at https://github.com/jbloomlab/SARS-CoV-2-BA.1_Spike_DMS_validations for BA.1 library validations and at https://github.com/jbloomlab/SARS-CoV-2-Delta_Spike_DMS_validations for Delta library validations.

2.6.13 Antibodies

LY-CoV1404 antibody was cloned and produced by GensScript. Variable domain sequences were taken from previously published antibody structure[102]. LY-CoV1404 variable regions were cloned with IgG1 heavy chain and human IgL2 constant regions, expressed in mammalian cells and purified using IgG-binding columns. S2 antibodies were produced as described previously in Zhou et al.[103]. 5-7 antibody was produced as described previously in Liu et al.[135]

2.6.14 Sera

267C and 279C sera are from Delta breakthrough cases that were previously vaccinated with two doses of Pfizer vaccine. 267C sample was collected from a male individual 28 days post symptom onset and 279C sample was collected from female individual 25 days post symptom onset. Before use, sera were heat inactivated at 56°C for 1 hour. These sera were part of prospective longitudinal Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) study from Washington State, USA.

2.6.15 Experimental Model and Subject Details

HEK-293T were acquired from ATCC (CRL3216), HEK-293T-ACE2 cells are described in Crawford et al.[70], generation and characterization of HEK-293T-ACE2-medium cells is described in Farrell et al.[134] (referred to ‘medium’ cells in the reference), generation of HEK-293T-rtTA cells is described below. All cells were grown in D10 media (Dulbecco’s Modified

Eagle Medium with 10% heat-inactivated fetal bovine serum, 2 mM l-glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin). For antibody selection experiments D10 was made with phenol-free DMEM (Corning DMEM With 4.5g/L Glucose, Sodium Pyruvate; Without L-Glutamine, Phenol Red from Fisher, Cat. No. MT17205CV). For experiments with HEK-293T-rtTA cells D10 was made with tetracycline-free FBS (Gemini Bio, Cat. No. 100-800) to avoid any expression of spike unless doxycycline is added.

To produce HEK-293T-rtTA expressing cells (used for storing deep mutational scanning libraries and required for TRE3G promoter activation) we first generated VSV-G-pseudotyped lentivirus carrying rtTA gene. To produce this virus we transfected 0.5 million HEK-293T cells with 0.25 µg of each helper plasmid (Gag/Pol, Tat1b, Rev1b), 0.25 µg of VSV-G expression plasmid and 1 µg of lentiviral backbone carrying plasmid into which rtTA has been cloned (plasmid map https://github.com/dms-vep/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/blob/main/library_design/reference_sequences/other_plasmid_maps_for_library_design/2850_pHAGE2_EF1a_TetOn_IRES_mCherry.gb). Note this lentiviral backbone is self-inactivating and therefore would not be produced and packaged into lentiviral particles upon helper plasmid transfection. 48 hours after transfection virus-containing cell supernatant was collected and filtered through a surfactant-free cellulose acetate 0.45 µm syringe filter. 5µl of the virus was used to infect 0.5 million low passage HEK-293T cells. 48 hours post infection single cell clones were sorted into a 96-well plate using BD Aria II cell sorter with 610/20 filter in PE-Texas Red channel. Single clones were expanded and tested for ability to produce high virus titers. High virus titers are essential for performing deep mutational scanning experiments on a practical experimental scale and we found that individual cell clones can vary significantly in the virus titers they can produce. Clonal cell population producing the highest virus titers was selected for expansion, frozen down in 10% DMSO and 20% FBS D10 media, and stored long-term in liquid nitrogen freezer.

2.7 Computational Analysis

2.7.1 Overview of data analysis pipeline

To analyze the deep mutational scanning data, we created a modular analysis pipeline. At the core of this pipeline is a set of common steps that are expected to be shared across analysis of many different datasets. We implemented this set of common steps in a standalone GitHub repository named dms-vep-pipeline that is publicly available at https://github.com/dms-vep/dms-vep-pipeline and is designed to be included in project-specific analyses as a git submodule. The dms-vep-pipeline consists of a series of Snakemake[136] rules that run Python scripts or Jupyter notebooks, and specifies a conda environment that provides details on the software used for the analysis. For this paper, we used version 1.01 of the dms-vep-pipeline.

For each specific project (in this case, deep mutational scanning of the BA.1 and Delta spikes) we created a separate GitHub repository that included dms-vep-pipeline as a submodule. The repository for BA.1 is publicly available at https://github.com/dms-vep/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs and the repository for Delta is at https://github.com/dms-vep/SARS-CoV-2_Delta_spike_DMS_REGN10933 Note how each repository has a configuration file (the config.yaml file), project-specific input data (the data subdirectory), and a top-level Snakemake file (the Snakefile) that runs the analysis. The output of running the pipeline is placed in a results subdirectory, although only key results files are tracked in the GitHub repository since many of them are very large. The pipeline also generates HTML rendering of the key analysis notebooks and result plots, which are available at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs for BA.1 and https://dms-vep.github.io/SARS-CoV-2_Delta_spike_DMS_REGN10933 for Delta. Looking at these websites is the easiest way to understand the analysis. Note that many of the plots are

interactive charts created with Altair[137], and we encourage readers to use the interactive features to better explore the data.

2.7.2 Analysis of PacBio data to link barcodes to spike mutations

To link each barcode to its spike variant, we used alignparse[138] to process the PacBio CCSs to determine the barcode and spike mutations for each CCS.

We performed several quality control steps. First we examined the synonymous tags introduced at the end of each amplicon during the library preparation to identify CCSs with discordant tags indicative of strand exchange during library preparation: typically <2% of CCSs had discordant tags, indicating a low rate of strand exchange. CCSs with identified strand exchange or out of frame indels were discarded. Next, we computed the empirical accuracy of the CCSs by examining how often CCSs with the same barcode reported the same spike sequence (the exact method used to compute the empirical accuracy is implemented here: https://jbloomlab.github.io/alignparse/alignparse.consensus.html#alignparse.consensus.empirical_accuracy). The empirical accuracies were between 0.65 and 0.75, indicating that fraction of CCSs correctly report the actual mutations. The inaccuracies are due to a combination of sequencing errors, reverse transcription errors, PCR strand exchange, and occasional actual association of the same barcode with different variants in different cells (which can especially arise if the complexity of the initial virus library integrated into cells at single copy is not much higher than the complexity of the final cell library).

We then built consensus sequences for each barcode with at least three CCSs, using the method implemented at https://jbloomlab.github.io/alignparse/alignparse.consensus.html#alignparse.consensus.simple_mutconsensus with max_minor_sub_frac and max_minor_indel_frac both set to 0.2. This approach of requiring multiple concordant CCSs to call a consensus is expected to lead to

higher accuracy in the final barcode / spike variant linking, and will generally discard barcodes that are not uniquely linked to a single spike variant.

Files containing the final barcode / variant lookup tables and the analysis notebooks with resulting quality control plots are linked to the main HTML pages in the documentation for the BA.1 and Delta experiments as provided in the Processed data section below.

2.7.3 Analysis of Illumina data to count barcodes for each variant in each experiment

For each experiment, we processed the Illumina barcode sequencing with the parser implemented at https://jbloomlab.github.io/dms_variants/dms_variants.illuminabarcodeparser.html to determine the counts of each variant in each condition. Barcoded variants were only retained for subsequent analysis if their “pre-selection” counts (no-antibody selection for antibody escape experiments, VSV-G-pseudotyped infections for functional selections) met some minimal count threshold specified in the config.yaml file of the GitHub repos for the BA.1 and Delta spikes. This thresholding removes variants that are expected to have substantial noise due to low counts. Note that a caveat that should be kept in mind is that the actual key bottleneck is expected to usually occur at the stage of infection with the virus library rather than sequencing, since the barcodes are generally sequenced to a depth that greatly exceeds the complexity of the libraries used for the infections. Therefore, although variants with low counts are expected to have more noise, the counts do not enable a quantitative estimate of the actual bottleneck size experienced by each variant.

Files containing the barcode counts and the analysis notebook with resulting quality control plots are linked to the main HTML pages in the documentation for the BA.1 and Delta experiments as provided in the Processed data section below.

2.7.4 Computing functional effects of mutations in deep mutational scanning

To estimate the functional effects of individual mutations, we first computed functional scores for each variant from the counts in the VSV-G-pseudotyped library infection (which should not impose any selection on the spike) versus the spike-pseudotyped library infected into ACE2 expressing target cells. The functional score for variant *v* is defined as $log_2\left(\left[n^v{}_{post} \,/\, n^{wt}{}_{post}\right]/\left[n^v{}_{pre} \,/\, n^{wt}{}_{pre}\right]\right)$ where $n^v{}_{post}$ is the count of variant *v* in the post-selection (spike-pseudotyped) infection, $n^v{}_{pre}$ is the count of variant *v* in the pre-selection (VSV-G-pseudotyped) infection, and $n^{wt}{}_{post}$ and $n^{wt}{}_{post}$ are the counts of all unmutated (wildtype) variants in each condition. Negative functional scores indicate a spike variant is worse at mediating infection than the unmutated spike, and positive functional scores indicate a variant is better at mediating infection than the unmutated spike. The distributions of these functional scores are plotted in Figure 2.5.

To deconvolve the functional scores for all spike variants (which often contain multiple mutations) into estimates of the effects of individual amino-acid mutations, we fit global epistasis models[76,77] to the variant functional scores, using the models implemented at https://jbloomlab.github.io/dms_variants/dms_variants.globalepistasis.html with the Gaussian likelihood function. This fitting estimates how each mutation affects an underlying latent phenotype, as well as the shape of the global epistasis function relating the latent phenotype to the observed functional score. We also then re-transform the inferred effect of each mutation on this latent phenotype through the global epistasis function to estimate its effect on the observed phenotype. This approach provides a way to deconvolve the information in the multi-mutant variants to more accurately estimate the effects of mutations under the assumptions of a global epistasis model[77]. For final reporting, we then took the average (median) of the estimated functional effect of each mutation across all the replicates and libraries for each different spike (BA.1 or Delta). In Figure 2.14 we report the functional effects on the observed (rather than

latent) phenotype, as that is a more relevant measure of its expected impact on spike-mediated infection.

Files containing the effects of mutations on both the latent and observed phenotypes for both individual replicates/libraries and averages across them, the analysis notebooks with relevant quality control plots, and interactive plots summarizing the final estimates are linked to the main HTML pages in the documentation for the BA.1 and Delta experiments as provided in the Processed data section below.

For the figures and the default view of the interactive plots described in the Processed data section below, we only include mutations seen in at least three distinct variants (averaged across libraries).

2.7.5 Computing antibody escape by mutations

For the antibody selections, we computed the non-neutralized fraction (probability of escape) $p_v(c)$ for each variant v at each antibody concentration c for a given antibody as $p_v(c) = (n_v{}^c/S^c)/(n_v{}^0/S^0)$ where $n_v{}^c$ is the count of variant v at antibody concentration c, $n_v{}^0$ is the count of variant v in the no-antibody control, S^c is the total counts of the neutralization standard at antibody concentration c, and S^0 is the total concentration of the neutralization standard in the no-antibody control. These values should in principle fall between 0 (variant is totally neutralized by antibody) and 1 (variant is not neutralized by antibody), and in practice we clip any values measured as >1 to a value of 1.

To deconvolve mutation-level escape values from the measured pvc values for the variants (which often contain multiple mutations) at multiple concentrations, we used the approach implemented in polyclonal software package (https://jbloomlab.github.io/polyclonal/)[78], constraining the fits to a single epitope (since we are only analyzing monoclonal antibodies). This analysis yields a mutation-level escape score for each observed variant (the m,e values in the nomenclature of the polyclonal package) which will be zero for mutations that have no effect

on antibody escape, and >0 for mutations that mediate antibody escape. These are the values plotted in the heat maps shown in the antibody escape figures; the line plots show site-level summaries of these values (eg, the sum of the escape values at each site). Note that the polyclonal models[78] can use the escape values inferred from the deep mutational scanning to predict the non-neutralized fraction for arbitrary mutants, and we correlated those measurements with the IC50 values measured by standard neutralization assays in the antibody escape figures. For all antibodies we had replicate measurements (multiple libraries, and in some cases technical replicates of the same library), and the final reported values are the average (median) across these replicates.

Files containing the escape values for both individual replicates/libraries and averages across them, the analysis notebooks with relevant quality control plots, and interactive plots summarizing the final estimates are linked to the main HTML pages in the documentation for the BA.1 and Delta experiments as provided in the Processed data section below.

For the figures in this paper and the default view of the interactive plots described in the Processed data section below, we only include mutations seen in at least three distinct variants (averaged across libraries), and with a functional effect of at least -1.38 (for Omicron BA.1) or at least -1.47 (for Delta).

2.7.6 Comparison of deep mutational scanning data to enrichment of mutations during actual human SARS-CoV-2 evolution

In Figure 2.14, we compare the functional effects of mutations measured in deep mutational scanning to the enrichment of mutations in actual human SARS-CoV-2 evolution compared that expected given the underlying viral mutation rate. The computation of the "expected" versus actual observed mutation counts is performed by the code at https://github.com/jbloomlab/SARS2-mut-rates/tree/25204ea0c868c4c86d0cc16bd5f203b1b0607868

The approach for calculating the enrichment of observed versus expected mutations is inspired by the strategy used in Neher[139]. Briefly, we downloaded the pre-built UShER tree of all publicly available SARS-CoV-2 sequences[97] as of Sept-26-2022 from http://hgdownload.soe.ucsc.edu/goldenPath/wuhCor1/UShER_SARS-CoV-2/2022/09/26/public-2022-09-26.all.masked.nextclade.pangolin.pb.gz. We then used matUtils[97] to separate the tree by the pre-annotated Nextstrain clades, and extract the number of unique occurrences of each nucleotide mutation along the tree branches for each clade. Note that we are counting unique occurrences of each mutation along the tree (number of branches with a mutation), not the number of observations of the mutation in the final set of aligned mutations. This is an important distinction, as a mutation could be common in the final aligned sequences even if it only had one or a few occurrences. In counting these occurrences, we ignored mutations on any branch with more than four nucleotide mutations, or more than one nucleotide mutation to either the Wuhan-Hu-1 reference sequence or the founder sequence for that clade, as taken from Neher[139]. The reason for excluding mutations on branches with high numbers of mutations as these can often indicate problematic sequences; the reason for excluding mutations on branches with multiple reversions to reference is that sometimes sequences with missing coverage have identities automatically called to reference. After tabulating these mutation counts, we ignored any mutations that are reversions from the clade founder to the Wuhan-Hu-1 reference or reversions of these, a set of sites known to be subject to sequencing errors[140], or were at sites with unusually high mutation rates (see sites manually specified in the config.yaml file at the GitHub repo listed above). Overall, this process provided a list of the counts of nucleotide mutations for each Nextstrain clade after filtering branches and mutations/sites expected to be affected by various possible errors.

We then calculated the relative rates of each type of the 12 possible nucleotide mutations at four-fold synonymous third codon positions by counting the mutation at such positions in each Nextstrain clade, and then normalizing the fraction of these positions that had

each nucleotide identity at each site, retaining only Nextstrain clades with at least 5000 such mutation counts[109]. We then calculated for each clade the expected number of counts of each nucleotide mutation based on these rates and the actual counts at the four-fold synonymous sites. The result of these calculations is how many times we expect to observe each nucleotide mutation if all sites are under the same constraint as the four-fold synonymous sites. Next, we aggregated the expected and observed counts across all Nextstrain clades, and also across all nucleotide mutations that encoded for the same amino-acid mutation to get expected and observed counts for amino-acid mutations, and only retained amino-acid mutations with at least 20 expected counts. Figure 2.14D plots the log base 2 of the expected versus observed counts after adding a pseudocount of 0.5.

We compared these counts to deep mutational scanning data from our study and several other spike deep mutational scanning studies as described in the main text. The tabulated deep mutational scanning measurements and the enrichments of mutations in actual SARS-CoV-2 evolution are at https://github.com/dms-vep/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/blob/main/results/compare_muteffects/all_natural_enrichment_vs_dms.csv

2.7.7 Processed data

The key results from the analysis are stored in the results subdirectory of the GitHub repos for BA.1 (https://github.com/dms-vep/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs) and Delta (https://github.com/dms-vep/SARS-CoV-2_Delta_spike_DMS_REGN10933). The easiest way to navigate these results are via the HTML documentation at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs and https://dms-vep.github.io/SARS-CoV-2_Delta_spike_DMS_REGN10933 These pages contain links to the key data files, as well as interactive heat maps of the functional effects of mutations and the effects of mutations on antibody escape. Note that these plots are interactive, and allow you to

filter by certain regions of the protein, the number of variants in which a mutation is seen, the maximum magnitude of an effect at a given site, and other relevant parameters.

In the final output files, mutations are numbered in reference-based (Wuhan-Hu-1) spike numbering. The GitHub repos contain files that convert sequential numbering of the BA.1 and Delta spike to reference based numbering.

2.7.8 Raw sequencing data

The raw PacBio and Illumina sequencing data have been deposited on the NCBI's Sequence Read Archive with BioProject number PRJNA888402 for the Omicron BA.1 data and PRJNA889020 for the Delta data. The PacBio sequencing linking variants to barcodes can be found under BioSample accessions SAMN31220980 for Omicron BA.1 and SAMN31230634 for Delta. The Illumina barcode sequencing can be found under BioSample accessions SAMN31216920 for Omicron BA.1 and SAMN31230628 for Delta.

2.8 Acknowledgments

We thank Michael Emerman for scientific advice and Kenneth A Matreyek for providing HEK-293T-ACE2-medium cell line. This work was supported in part by the NIH/NIAID under grant R01AI141707 and contract 75N93021C00015 to JDB. This work was supported in part by a Gates Foundation grant INV-004949 to JDB. JDB is an Investigator of the Howard Hughes Medical Institute. KHDC was supported by NIH/NIAID grant F30AI149928. BD was funded in part by EMBO under grant ALTF 81-2020. CER was supported by the Viral Pathogenesis and Evolution training grant T32 AI083203. TCY was supported by the CMB training grant T32 GM007270 and the NSF graduate research fellowship DGE-2140004. DRB was supported by NIAID grant UM1 AI44462 and Bill and Melinda Gates Foundation grants OPP1170236 and INV-004923.

2.8.1 Ethics statement

Delta breakthrough sera were collected after written and informed consent. The breakthrough sera were part of the prospective longitudinal Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) study from Washington State, USA, which was approved by University of Washington Institutional Review Board (protocol #STUDY00000959).

2.8.2 Competing Interests

JDB is on the scientific advisory board of Apriori Bio, Aerium Therapeutics, Invivyd, the Vaccine Company, and Oncorus, and has recently consulted on topics related to viral evolution for Moderna and Merck. JDB, KHDC, and CER receive royalty payments as inventors on Fred Hutch licensed patents related to viral deep mutational scanning. JDB, KHDC, CER and BD are inventors on a pending patent application relating to the viral deep mutational scanning system described in this paper. RB is a consultant for IAVI, Adagio, Adimab, Mabloc, VosBio, Nonigenex, and Radiant. DDH is a co-founder of TaiMed Biologics and RenBio, consultant to WuXi Biologics and Brii Biosciences, and board director for Vicarious Surgical. HYC consulted with Ellume, Pfizer, The Bill and Melinda Gates Foundation, GlaxoSmithKline, and Merck. HYC received research funding from Gates Ventures, Sanofi Pasteur, and support and reagents from Ellume and Cepheid outside of the submitted work. The other authors declare no competing interests.

Chapter III: Mapping the neutralizing specificity of human anti-HIV serum by deep mutational scanning

A version of this chapter has been published as:

Radford, C. E., Schommers, P., Gieselmann, L., Crawford, K. H., Dadonaite, B., Yu, T. C., ... & Bloom, J. D. (2023). Mapping the neutralizing specificity of human anti-HIV serum by deep mutational scanning. *bioRxiv*, 2023-03.

This chapter describes my use of the lentivirus deep mutational scanning system to map HIV Envelope escape from neutralization by antibodies and human sera. By enabling the measurement of the effects of combinations of mutations, the lentivirus deep mutational scanning system enabled us to map neutralization activity of polyclonal sera at multiple epitopes simultaneously.

3.1 Abstract

Understanding the specificities of human serum antibodies that broadly neutralize HIV can inform prevention and treatment strategies. Here we describe a deep mutational scanning system that can measure the effects of combinations of mutations to HIV envelope (Env) on neutralization by antibodies and polyclonal serum. We first show that this system can accurately map how all functionally tolerated mutations to Env affect neutralization by monoclonal antibodies. We then comprehensively map Env mutations that affect neutralization by a set of human polyclonal sera known to target the CD4-binding site that neutralize diverse strains of HIV. The neutralizing activities of these sera target different epitopes, with most sera having specificities reminiscent of individual characterized monoclonal antibodies, but one sera targeting two epitopes within the CD4 binding site. Mapping the specificity of the neutralizing

activity in polyclonal human serum will aid in assessing anti-HIV immune responses to inform prevention strategies.

3.2 Introduction

Efforts to create a HIV vaccine have been stymied in part by rapid and continuing diversification of the virus's envelope (Env) protein[30,38]. However, some individuals with HIV do naturally develop polyclonal serum antibody responses to Env that broadly neutralize many viral strains[4,36,37]. Much progress has been made characterizing individual broadly neutralizing antibodies. However, individual antibodies do not always recapitulate the neutralizing activity of the serum of the individuals from whom they were isolated[141–144].

Mapping the specificity of polyclonal neutralizing serum antibodies is more difficult than characterizing individual monoclonal antibodies. One important advance has been the development of electron microscopy-based polyclonal epitope mapping (emPEM) methods to visualize how multiple different serum antibody Fabs bind to Env[64–66]. However, this approach characterizes binding rather than neutralizing specificity, and one major finding from emPEM is that many serum antibodies bind non-neutralizing epitopes[64–66,51]. Fingerprinting approaches can define neutralizing epitopes, but does not provide mutation-level specificity and requires making measurements for large virus panels[145,146]. Deep mutational scanning can map Env mutations that escape antibody neutralization[46,49–51,147]. However, existing HIV deep mutational scanning work has used approaches that are only able to look at effects of individual mutations, which is a limitation when trying to map polyclonal serum antibodies that may target multiple epitopes[51].

Precisely mapping neutralizing specificities and escape mutations is especially challenging for antibodies that target the CD4-binding site. Such antibodies recognize conserved Env residues while typically avoiding steric clashes with glycans rather than

depending on them for neutralization, unlike antibodies targeting other epitopes such as the V1/V2 loops or V3 loop[4,36]. As a result, CD4-binding site targeting antibodies can have near pan-HIV neutralization breadth and high potency despite sequence and glycan heterogeneity across strains of HIV[4,36,37], and are therefore promising candidates for treatment and prophylaxis strategies[37,148,149]. But the higher conservation of their epitopes can also make it more difficult to map escape mutations for such antibodies[147].

Here we use an improved deep mutational scanning system to measure how mutations affect neutralization of Env by human anti-HIV sera that target the CD4-binding site. This new system can measure the effects of combinations of mutations, enabling quantitative deconvolution of how mutations mediate escape at distinct antibody epitopes[78]. We find that several sera have neutralizing activities that resemble monoclonal antibodies, but one sera has neutralizing activity targeting two distinct epitopes. These maps shed light on the specificity of human serum that can broadly neutralize many HIV strains. In addition, the method we employ could be used in the future to evaluate and compare the neutralizing specificities of anti-HIV sera elicited by different vaccine regimens.

3.3 Results

3.3.1 Single-round replicative lentivirus deep mutational scanning platform for HIV Env

We recently described a deep mutational scanning platform based on a single-round replicative lentivirus that does not encode any viral genes except for the viral entry protein[150], which in our current study is Env. This platform enables creation of large libraries of single-round replicative lentiviruses with a genotype-phenotype link between barcodes in the lentivirus genomes and the entry proteins on the surfaces of virions (Figure 3.1A,B). Key aspects of this platform include encoding viral entry protein mutants in lentivirus genomes with random

nucleotide barcodes and using a lentivirus genome with a full 3' LTR that can be reactivated after infection[67,92,150] (Figure 3.1A). Creation of the mutant libraries involves a two-step process of first integrating lentivirus genomes into cells at just one copy per cell, and then generating mutant virus libraries with a genotype-phenotype link (Figure 3.1B). PacBio sequencing is used to link Env mutants with their nucleotide barcodes, and later experiments use short read Illumina sequencing of the barcodes to measure mutant frequencies.

Figure 3.1 Lentivirus platform for deep mutational scanning

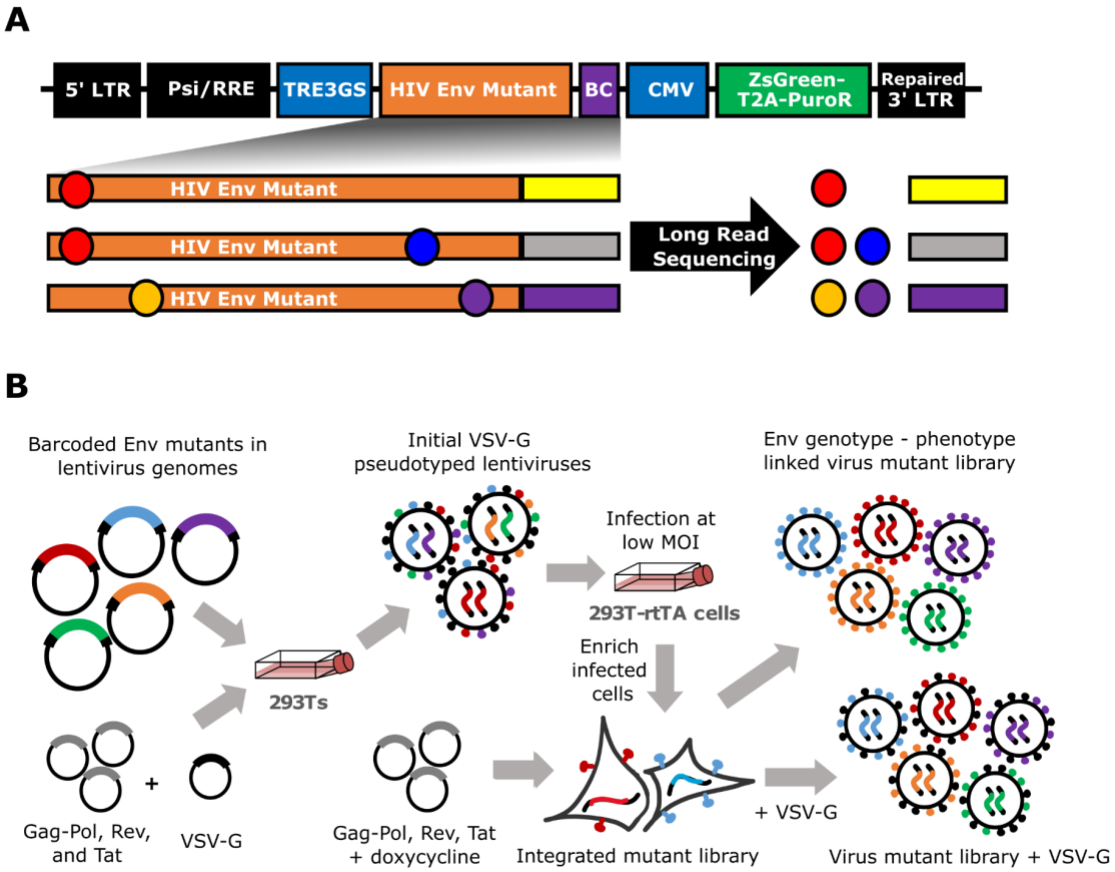


Figure 3.1: (A) The lentivirus genome used for deep mutational scanning. The genome contains the full 5´ and 3´ LTR sequences, including the U3 sequence usually deleted in the 3´ LTR. Env is under control of an inducible TRE3G promoter and followed by a 16N random nucleotide barcode. A CMV promoter drives ZsGreen and puromycin resistance (PuR) expression. (B) Approach for generating genotype-phenotype linked variant libraries. Lentivirus genomes carrying barcoded Env mutants are transfected into 293T cells alongside plasmids expressing the lentiviral proteins necessary for creating single-cycle infectious virions and VSV-G. The

resulting VSV-G pseudotyped viruses are used to infect 293T-rtTA cells at a low multiplicity of infection, such that most infected cells receive just one viral genome. Infected cells are enriched via puromycin selection, and genotype-phenotype linked Env-expressing virus variant libraries are generated by inducing Env expression with doxycycline and transfecting plasmids encoding the lentivirus genes. The virus variant libraries are also generated separately with VSV-G, and these VSV-G pseudotyped viruses can infect cells regardless of whether or not they have a functional Env and so can be used to readout the library composition.

3.3.2 Env mutant library design and generation

Our libraries used Env from the transmitted/founder virus BF520.W14M.C2[151,152] (referred to hereafter as BF520). We chose this Env for several reasons. First, since it is from a transmitted/founder virus, it is particularly relevant for antibody neutralization studies[141,152,153]. Second, the BF520 Env yielded high titers in our lentiviral deep mutational scanning platform (Figure 3.2). Third, our lab has previously performed deep mutational scanning of the BF520 Env using full-length replicative HIV libraries in a prior system that did not enable analysis of multiply mutated Env mutants[48,49], thereby providing comparator data to benchmark the current study.

Figure 3.2 Titers of Env or VSV-G pseudotyped lentiviruses on TZM-bl cells

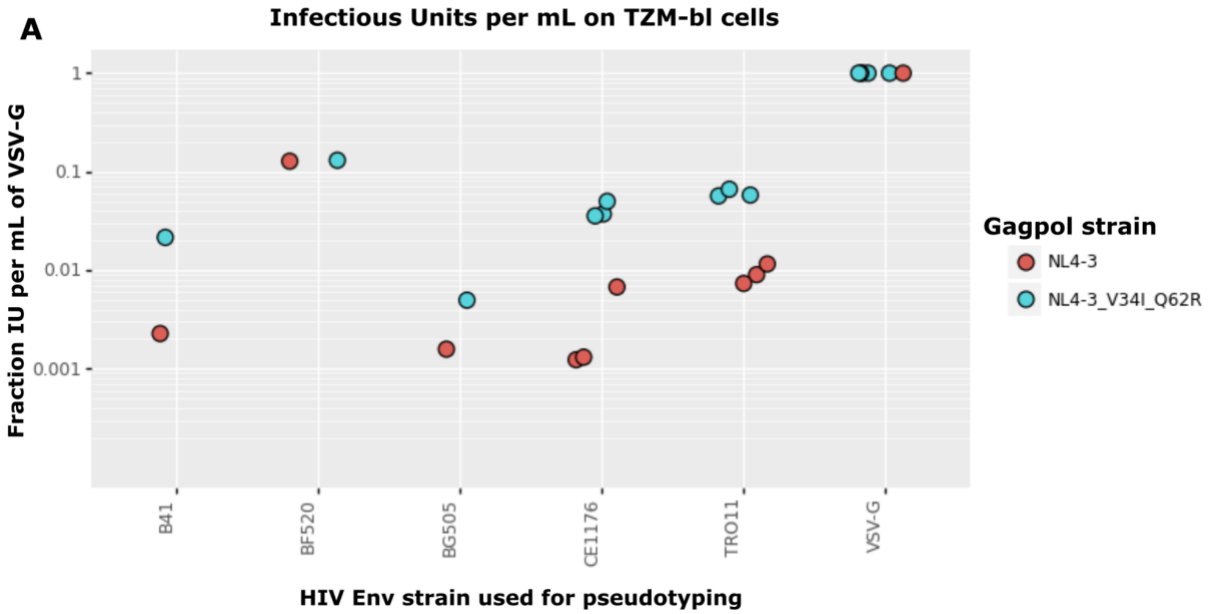


Figure 3.2: Lentiviruses were produced by transfecting 293T cells with the same ZsGreen reporter lentiviral backbone, Rev, and Tat expressing plasmids for each condition, along with an HIV Env from the indicated viral strain or a VSV-G expressing plasmid, and either a NL4-3 or NL4-3 V34I Q26R Gag-Pol expressing plasmid as indicated. Data are from different virus preparations and titering dates. The infectious units per mL on TZM-bl cells were normalized to VSV-G infectious units per mL by dividing each condition's infectious units per mL by that of the VSV-G pseudotyped virus with the same Gag-Pol that was produced and titered on the same dates, performing this normalization to correct for batch effects. The titers for the VSV-G pseudotyped viruses ranged from ~1.5-35 million infectious units per mL. Gag-Pol mutations V34I and Q62R were made based on previous studies that showed these mutations can rescue Env incorporation deficiencies[154,155].

A goal of our experiments is to map escape from polyclonal serum antibodies as well as monoclonal antibodies. Since polyclonal serum is composed of multiple antibodies that can target different epitopes[156,157,10], mapping escape from serum requires libraries that contain Envs with mutations in multiple epitopes[78]. However, about half of all possible amino-acid mutations to proteins are highly deleterious,[47,48,158,159] so a library of multiply mutated Envs with random mutations would contain a high fraction of non-functional proteins. Therefore, we designed the libraries to exclude most highly deleterious mutations. We also mutagenized only the Env ectodomain (and not the signal peptide, transmembrane domain, and cytoplasmic tail), since neutralizing antibodies always bind the ectodomain.

To design libraries containing mostly functional Env mutants, we drew on two sources of information. The first source of information was prior deep mutational scanning data for BF520 Env generated using full-length replicative HIV virions in a system that could only measure the average effect of mutations across different genetic backgrounds[48,49]. We used data from this prior deep mutational scanning to identify well-tolerated mutations (Figure 3.3A, left panel). The second source of information was an alignment of group M HIV-1 sequences[160], which we used to identify any mutations relative to BF520 present more than once in natural sequences (Figure 3.3A, middle panel). Our library design included the 7110 amino-acid mutations in the BF520 ectodomain that were either tolerated in the prior deep mutational scanning or present multiple times in the natural sequence alignment (Figure 3.3A, right panel). We created these Env

libraries using a previously described PCR mutagenesis approach modified to target these specific mutations[45,150].

We generated two independent Env mutant libraries to perform the deep mutational scanning in biological duplicate. PacBio sequencing showed that each library had ~2.5 nonsynonymous mutations per Env mutant, which are linked via the barcode and so can be evaluated in combination (Figure 3.1A). There was a low frequency of synonymous mutations, stop codons, and in-frame deletions (Figure 3.3B). Overall, ~84% of the mutations were among the 7,110 mutations targeted in our library design. Each library contained ~40,000 barcoded mutants, and together the two libraries sampled ~97% of the targeted mutations (Figure 3.3C).

Figure 3.3 Mutant library design and functional effects of mutations

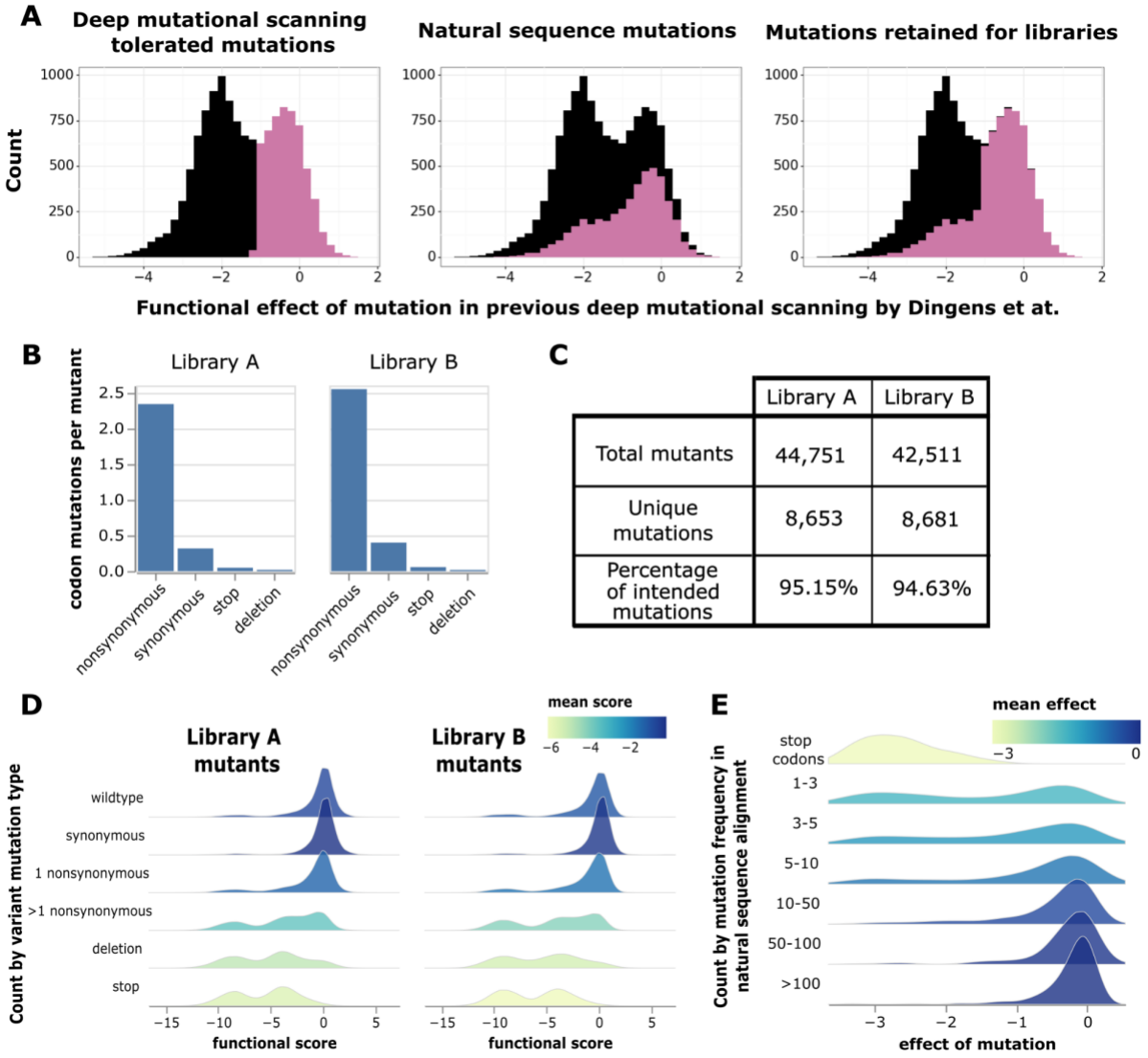


Figure 3.3: (A) Choice of targeted mutations based on measured effects in prior deep mutational scanning[48] and occurrences in natural HIV sequences. The distributions of previously measured mutation effects are shown for all mutations to BF520 (black) with highlighting of subsets of mutations (purple). From left to right, highlighted are mutations well tolerated in the prior deep mutational scanning[48], mutations found multiple times in natural sequences, and the union of these two sets. Mutations in the union of the two sets were designed into our libraries. (B) Average codon mutations per Env mutant, separated by type of codon mutation. (C) Total number of barcoded Env mutants in each library, along with the numbers of unique mutations and percentage of the intended mutations present across these mutants. (D) Distributions of functional scores measured in our deep mutational scanning across Env mutants, separated by the types of codon mutations found in the mutants. Negative functional scores indicate impaired Env-mediated infection relative to unmutated BF520 Env. Histograms are colored by mean functional score. (E) Distributions of mutation effects versus how often that mutation is found in natural sequences. The distribution of stop codon effects is also shown.

3.3.3 Effects of mutations on Env-mediated viral entry

To measure how mutations affected the ability of Env to mediate viral infection in cell culture, we generated libraries pseudotyped with just the Env mutants or also pseudotyped with VSV-G (Figure 3.1B). We then used these libraries to infect TZM-bl cells[161–163], which express Env's primary receptor (CD4) and co-receptors (CCR5 and CXCR4). We sequenced the barcodes of virions that infected cells in each condition: all virions are expected to infect cells when VSV-G is present, but only virions with functional Envs will infect cells in the absence of VSV-G. Each barcoded Env variant was assigned a functional score calculated as the log of the ratio of the frequency of that variant (relative to unmutated BF520 Env) in the Env versus VSV-G mediated infections. Negative functional scores indicate an Env mutant is worse at infecting cells than unmutated BF520 Env, while positive functional scores indicate it is better at infecting cells.

As expected, Env mutants with only synonymous mutations have "wildtype-like" functional scores of near zero, while mutants with stop codons generally have highly negative functional scores (Figure 3.3D). Most mutants in the libraries with only one nonsynonymous mutation have functional scores close to zero, suggesting our library design largely incorporated functionally tolerated mutations as intended. Env mutants with multiple nonsynonymous mutations more often have substantially negative functional scores, as expected from the accumulation of multiple sometimes deleterious mutations (Figure 3.3D).

To estimate the effect of each individual mutation on viral entry, we fit global epistasis models to the functional scores[76,77]. The effects of mutations on Env-mediated viral entry are shown at https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/muteffects_observed_heatmap.html in an interactive heatmap. Mutations found more often among natural sequences tend to have more favorable effects in our experiments than mutations only found rarely among natural

sequences (Figure 3.3E), suggesting mutations that are favorable for viral entry in our experiments are generally also favorable during natural HIV evolution.

3.3.4 Accurate mapping of effects of Env mutations on antibody neutralization

We next used the deep mutational scanning platform to map how Env mutations affect antibody neutralization. As a first proof-of-principle, we mapped mutational escape from neutralization by the well-characterized broadly neutralizing antibody PGT151[164]. To directly quantify how mutations affected neutralization, we included a non-neutralized "standard" virus pseudotyped with VSV-G to enable conversion of relative sequencing counts to absolute neutralization measurements[150]. To estimate the effects of individual mutations from our library measurements (which include both singly and multiply mutated Envs), we fit a biophysical model where antibody neutralization at each epitope has a Hill-curve dependence on antibody concentration and mutations within a given epitope have additive effects on antibody affinity[78]. This model, which is implemented in the polyclonal software (https://jbloomlab.github.io/polyclonal/), utilizes information from both singly and multiply mutated Env variants under realistic assumptions about how mutations combine to escape antibody binding.

Our mapping showed that PGT151 is escaped by mutations in the fusion peptide or affecting N-linked glycans recognized by PGT151 (Figure 3.4A,B and interactive escape maps linked in figure legend). In particular, PGT151 is strongly escaped by any mutations knocking out the N611 glycan, specific mutations at the N637 glycan, mutations to positively charged residues at sites around the N637 glycan, mutations at sites 647 and 648, and mutations at sites 512 and 514 in the fusion peptide (Figure 3.4A,B). We also mapped lower magnitude escape at sites 537-543. All these mutations are in or near the binding footprint of PGT151[165] (Figure 3.4C).

Figure 3.4 Neutralization escape map for antibody PGT151

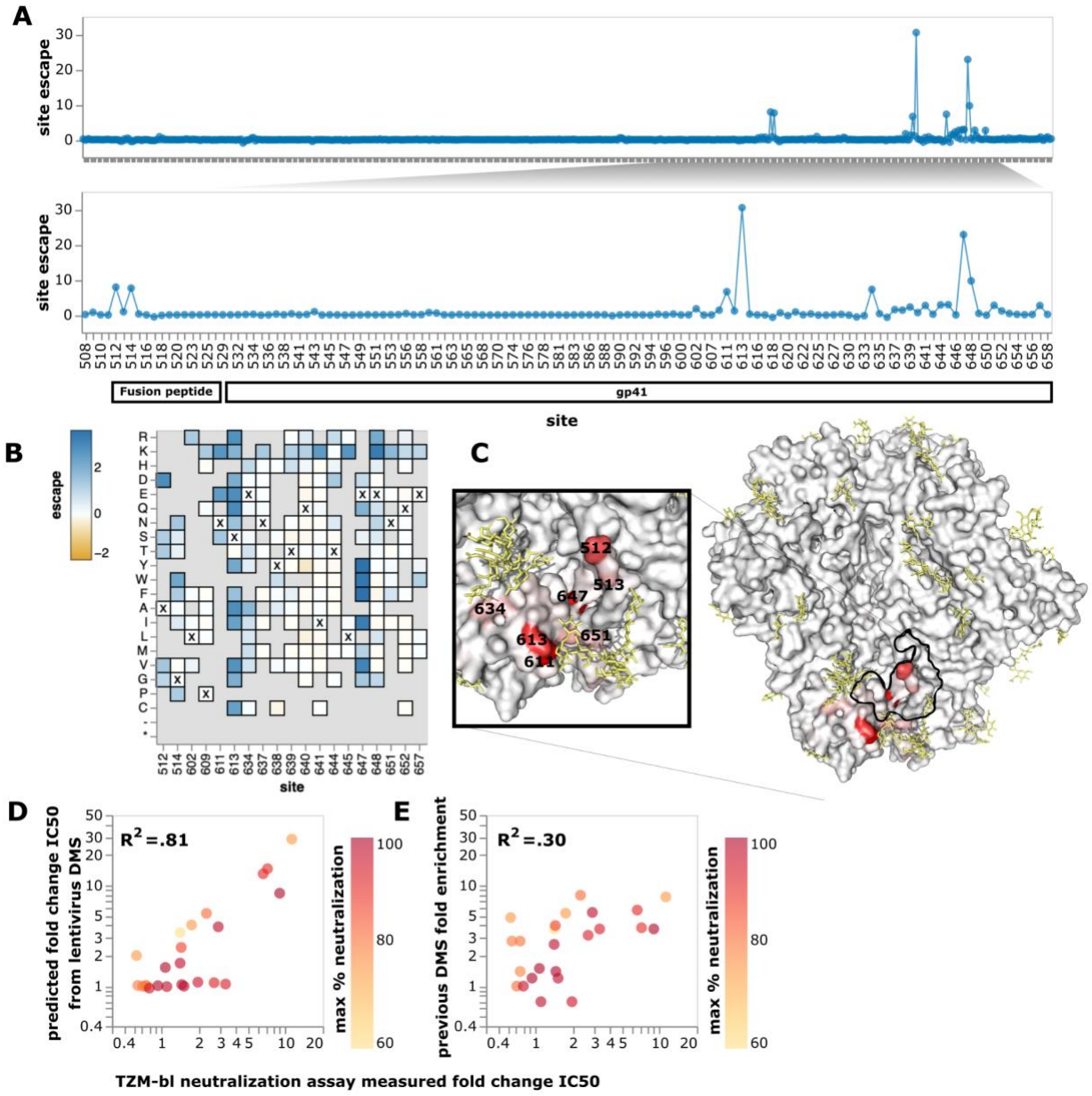


Figure 3.4: (A) The top panel shows PGT151 escape across all sites in the BF520 Env ectodomain, and the bottom panel zooms into key sites. The y-axis shows escape summed across all mutations at each site. (B) Heatmap of effects of individual mutations at key sites of escape. Residues marked with X are wildtype residues in BF520. Residues grayed out are not present in the variant libraries, typically because they are deleterious for Env function. See https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/PGT151_escape_plot.html for an interactive version of the site and mutation escape plots. (C) Site escape mapped onto a structure of PGT151-bound Env, with red indicating sites where mutations cause escape. Residues within 4 angstroms of antibody PGT151 in the structure are outlined in black. Glycans are colored yellow. This visualization was generated using the structure of BG505 ΔCT N332T (PDB 6MAR, antibody PGT151 removed). (D) Correlation of predicted fold-change in IC50 from the current deep mutational scanning versus fold change in IC50 as previously measured in

TZM-bl neutralization assays[49]. Mutations are colored by the max neutralization plateau observed for that variant in TZM-bl neutralization assays using PGT151. Pearson correlation coefficients were calculated on the fold-changes on a linear scale. Note the fold change IC50 predicted by deep mutational scanning is calculated using the full biophysical model fit to the data (see Methods). (E) Correlation of fold enrichment in prior HIV Env deep mutational scanning[49] versus fold-change in IC50 for the same set of single mutants.

The escape predicted by our deep mutational scanning was highly correlated with IC50s measured in previously performed TZM-bl neutralization assays[49] (Figure 3.4D). The correlation with neutralization IC50s was substantially better for our current deep mutational scanning than an earlier approach that used libraries of HIV virions in a system where it was not possible to measure multiple mutations or absolute neutralization[49] (Figure 3.4D,E).

3.3.5 Broadly neutralizing anti-HIV sera

We assembled a set of sera from individuals with HIV to test if we could map neutralizing specificities in a polyclonal context[147]. We chose sera based on their ability to broadly neutralize a global HIV panel[166] and potently neutralize BF520 pseudovirus. Based on these criteria, we chose four sera collected from individuals in Germany living with HIV (Figure 3.5A,B). Based on the f61 neutralization fingerprinting panel[145], these sera were predicted to be primarily VRC01-like, meaning they target the CD4-binding site (Figure 3.5C). Note that all the sera in our study target the CD4-binding site because we chose broad sera that neutralized BF520 (which is relatively resistant to V3 antibodies[141]); neutralizing human anti-HIV sera can target other epitopes[145,167,168]. Importantly, we used purified IgGs from these sera for our experiments, since antiretroviral drugs present in the sera could interfere with our lentiviral-based assays.

Figure 3.5 Broadly neutralizing human anti-HIV sera

A

	398F1	246F3	CNE55	CNE8	X2278	Tro11	BJOX2000	CH119	CE1176	CE0217	25710	X1632	BF520	MLV		
	A	AC	AE	AE	B	B	BC	BC	C	C	C	G	A	NA	Mean Neutr.	Breadth (%)
IDC#0508	42	50	41	36	85	71	399	71	229	127	94	120	28	>500	86	100
IDC#0513	136	281	140	102	75	85	277	174	271	>500	96	233	337	>500	152	92
IDC#0561	171	85	124	60	16	38	42	114	453	111	69	94	49	>500	83	100
IDF#033	25	36	69	72	52	28	179	207	62	63	66	188	60	>500	69	100

B

Patient ID	Country	Gender	Ethnicity	CDC Status	Viral load at blood draw	CD4 Count	On ART?	Virus Clade	Years since HIV-1 Diagnosis	Age (years)
IDC#0508	Germany	M	Middle Europe	A2	118000	450	No	D	8.91	37
IDC#0513	Germany	M	Middle Europe	NA	42	445	Yes	B	30.66	66
IDC#0561	Germany	M	Middle Europe	A2	5530	550	No	B	20.19	48
IDF#033	Germany	F	Central/South Africa	B3	19	313	Yes	D	20.38	53

C

	VRC01-like	b12-like	HJ16-like	8ANC195-like	PG9-like	PGT128-like	2F5-like	10E8-like	35O22-like	PGT151-like	Neut breadth (%)
IDC#0508	0.71	0.10			0.17					0.02	90%
IDC#0513	0.40			0.13	0.23			0.24			70%
IDC#0561	0.77		0.06				0.04		0.13		75%
IDF#033	0.33		0.14	0.04		0.32	0.11		0.05		60%

Figure 3.5: (A) Neutralization of a global HIV panel by each serum[166]. Values reported are IC50 in ug/mL for purified IgGs. (B) Clinical data related to the individual from whom each serum was collected. (C) f61 neutralization fingerprinting results for each sera[145].

3.3.6 Neutralization escape maps of serum IDC561 and its constituent antibody 1-18 are similar

We first analyzed serum IDC561, which was collected from the same individual from whom the broadly neutralizing antibody 1-18 was isolated[147]. The antibody was isolated from B cells from the same blood draw date as the serum we used, suggesting antibody 1-18 is likely

present in the serum. We mapped escape from neutralization by antibody 1-18 alongside the serum in order to compare the escape maps. It has been previously reported that 1-18 and purified IgGs from IDC561 display similar neutralization of a panel of viral strains and mutants, suggesting that neutralization by serum IDC561 is dominated by 1-18[147]. We therefore hypothesized that the neutralization escape map of serum IDC561 might resemble that of 1-18.

The maps for serum IDC561 and antibody 1-18 generally show neutralization escape at the same sites in Env, although the relative magnitude differs between the serum and antibody (Figure 3.6, and interactive escape maps linked in figure legend). In particular, both the serum and antibody are escaped by mutations around the V1/V2 loop, at β20/β21, and at the β23-V5-β24 structure (Figure 3.6A,B,C,D). Around the V1/V2 loop, the greatest escape from 1-18 is by mutations at site 198 in the middle of the N197 glycosylation motif (Figure 3.6A,B,D), and by mutations to sites 202, 203, and 206 (Figure 3.6A,B,D). IDC561 is also escaped by mutations at site 198, but mutations at sites 202 and 203 cause more escape for the serum than for 1-18, while there is less escape at site 206 for the serum than for 1-18 (Figure 3.6A,C,D). At β20/β21, mutations at sites 428-430 escape both 1-18 and IDC561, but the magnitude of this effect is lower for IDC561 than 1-18 (Figure 3.6A,B,C,D). At the β23-V5-β24 structure, mutations to sites 471, 474 and 476 escape 1-18, but only mutations at site 471 strongly escape IDC561 (Figure 3.6A,B,C,D).

The escape map for serum IDC561 was substantially more similar to that of antibody 1-18 than another CD4 binding site antibody, 3BNC117[169] , as well as the fusion peptide/gp120-gp41 interface-targeting antibody PGT151 (Figure 3.6E). This similarity suggests that antibody 1-18, which was isolated from the individual from which serum IDC561 was obtained, contributes substantially to overall neutralization by this serum as suggested by prior studies[147]. However, the fact that the serum IDC561 map does not entirely mirror that of 1-18 shows that other antibodies or members of the same clonal family also contribute to serum neutralization.

Figure 3.6 Neutralization escape map for antibody 1-18 and purified IgGs from IDC561

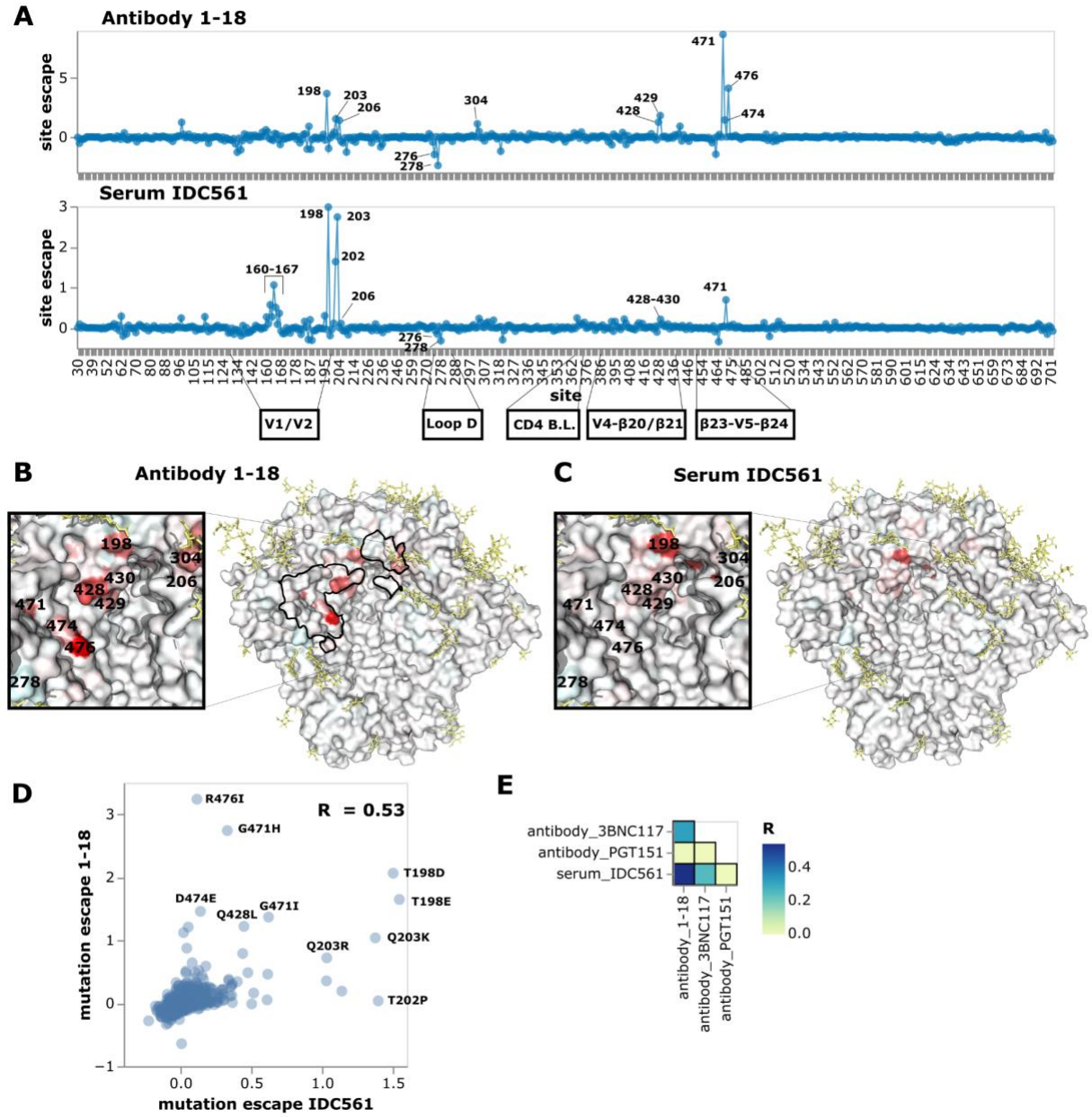


Figure 3.6: (A) Escape at all sites in the BF520 Env ectodomain from antibody 1-18 and serum IDC561. Positive values represent sites where mutations cause escape on average, while negative values represent sites where mutations enhance neutralization on average. See https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/1-18_escape_plot.html and https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/IDC561_escape_plot.html for interactive versions of the escape maps for 1-18 and IDC561, respectively.
(B) Site escape from 1-18 mapped onto a structure of 1-18-bound Env. Residues within 4 angstroms of antibody 1-18 in the structure are outlined in black. This visualization was generated using the structure of BG505.SOSIP.664 (PDB 6UDJ, antibodies 10-1074 and 1-18 removed)[147]. (C) Site escape from IDC561 mapped onto the same structure. (D) Scatter plot of

how mutations escape serum IDC561 versus antibody 1-18. (E) Correlations of how mutations escape IDC561 versus antibodies 1-18, 3BNC117, or PGT151.

3.3.7 Escape maps of other sera show diverse patterns of neutralization specificity

We next analyzed three more CD4-binding site targeting sera. The first of these sera, IDC513, was most escaped by mutations in loop D, similar to the well-characterized antibody 3BNC117 (Figure 3.7, and interactive escape maps linked in figure legend), although that antibody was not isolated from this individual. Both 3BNC117 and IDC513 are escaped by mutations in loop D, particularly at site 281 (Figure 3.7A-D). However, mutations at sites 276 and 278, which knock out the N276 glycan, enhance neutralization by both 3BNC117 and IDC513 (Figure 3.7A-D). These mutations also sensitize Env to neutralization by 1-18 and IDC561, but not to the same extent (Figure 3.6A). Mutations at sites 456, 459, and 471 in the β23-V5-β24 structure also escape both IDC513 and 3BNC117, and there is lower magnitude escape by mutations in and around the CD4 binding loop and other variable loops (Figure 3.7A-C). Overall the escape map for IDC513 correlates better with 3BNC117 than 1-18 (Figure 3.7D,E). Because 3BNC117 and serum IDC513 are from different individuals, neutralizing antibodies in serum IDC513 must have convergently evolved to target similar sites as antibody 3BNC117. Convergent evolution of broadly neutralizing HIV antibodies from the same heavy chain genes has been observed previously[169], although we do not know the genes encoding the neutralizing antibodies in serum IDC513. Note that efforts to induce similar antibody specificities form the basis of some vaccine strategies[170,171].

Figure 3.7 Neutralization escape maps for antibody 3BNC117 and purified IgGs from IDC513

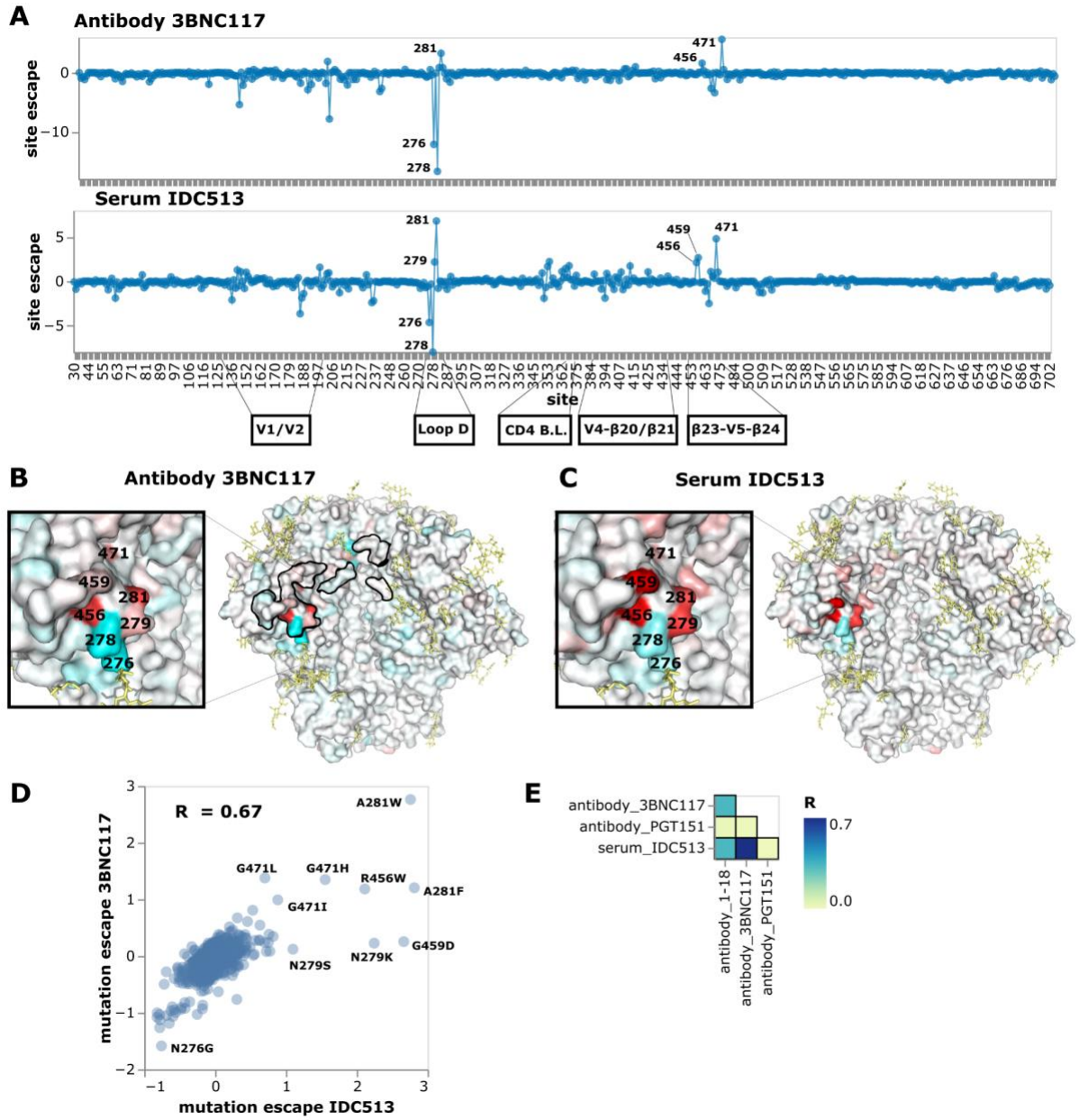


Figure 3.7: (A) Escape at all sites in BF520 Env ectodomain from antibody 3BNC117 and serum IDC513. See https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/3BNC117_escape_plot.html and https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/IDC513_escape_plot.html for interactive versions of the escape maps for 3BNC117 and IDC513, respectively. (B) Site escape from 3BNC117 mapped onto a structure of 3BNC117-bound Env. Residues within 4 angstroms of antibody 3BNC117 in the structure are outlined in black. This visualization was generated using the structure of BG505.SOSIP.664 along with antibody 3BNC117 (PDB 5V8M)[172]. (C) Site escape from IDC513 mapped onto the same structure. (D) Scatter plot of

how mutations escape serum IDC513 versus antibody 3BNC117. (E) Correlations of how mutations escape IDC513 versus antibodies 1-18, 3BNC117, or PGT151.

In contrast to IDC513 and IDC561, the escape map of IDF033 reveals a dependence on the N276 glycan for neutralization (Figure 3.8A,B, and interactive escape maps linked in the figure legend). Mutations at sites 276 and 278 that ablate the N276 glycan cause by far the greatest escape (Figure 3.8A,B). Other mutations in loop D, particularly at site 281, also more weakly escape from IDF033 (Figure 3.8A,B). At the β23-V5-β24 structure, mutations at sites 463 and 465 of the N463 glycosylation motif enhance neutralization by IDF033, but the mutation N463S causes escape by shifting the glycosylation motif to N461 (Figure 3.8A, Figure 3.9A). Other nearby sites also have mutation-specific effects (Figure 3.9A). For example at site S460, only some of the amino-acid changes cause escape (Figure 3.9A). Note that the neutralization fingerprinting panel (Figure 3.5C) suggests serum IDF033 also has some V3-targeting activity, but this is not apparent in our escape maps probably because BF520 has a relatively high baseline resistance to V3 targeting antibodies[141].

The escape map for the final serum, IDC508, revealed neutralization escape at two distinct antibody epitopes (Figure 3.8A,C,D, and interactive escape maps linked in figure legend). The existence of two epitopes was inferred by fitting the biophysical model[78] to the deep mutational scanning measurements and finding that escape in multiply mutated variants was best explained by mutations affecting antibody binding at two distinct regions. Note that identification of two separate epitopes is crucially enabled by the ability of our deep mutational scanning system to quantify escape by Envs with multiple mutations[78].

The first IDC508 epitope depends on the presence of the N276 glycan for neutralization and therefore is escaped by mutations at sites 276 and 278 as well as other mutations in loop D, similar to IDF033 (Figure 3.8A,C). Neutralization at this first epitope is also escaped by mutations at the β23-V5-β24 structure, also similar to IDF033 (Figure 3.8A,B,C, Figure 3.5B). The second IDC508 epitope mapped mainly to sites around the V1/V2 loop (Figure 3.8A,D).

Mutations at site 198 cause escape from neutralization at this second epitope, similar to 1-18 and IDC561 (Figure 3.6A,C,D, Figure 3.8A,D). Mutations at site 201, 202, and 203 and in the V2 loop at sites 160-167 also escape at the second epitope, again similar to IDC561 (Figure 3.6A, Figure 3.8A,D). Therefore, each of the two epitopes targeted by the neutralizing activity of IDC508 resembled the epitope targeted by another serum.

Figure 3.8 Neutralization escape maps for purified IgGs from IDF033 and IDC508

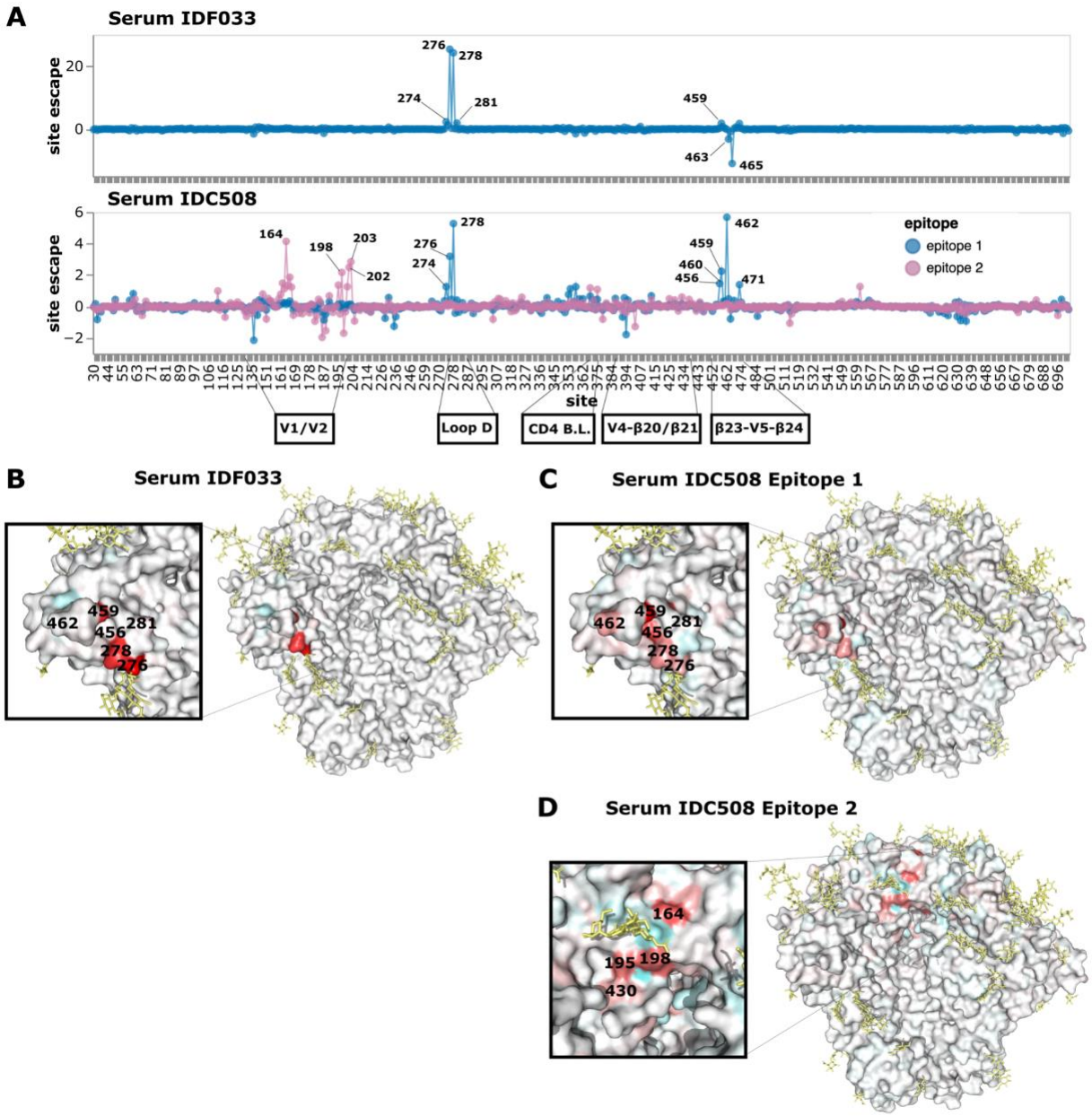


Figure 3.8: (A) Escape at all sites in BF520 Env ectodomain from serum IDF033 and serum IDC508. See https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/IDF033_escape_plot.html and https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/IDC508_escape_plot.html for interactive versions of the escape maps for IDF033 and IDC508, respectively. (B) Site escape from IDF033 mapped onto a structure of Env. This visualization was generated using the structure of BG505.SOSIP.664 (PDB 6UDJ, antibodies 10-1074 and 1-18 removed)[147]. (C) Site escape from the first IDC508 epitope mapped onto the same structure. (D) Site escape from the second IDC508 epitope mapped onto the same structure.

Figure 3.9 Zoomed in views of mutation-level escape at some key sites for IDF033 and IDC508

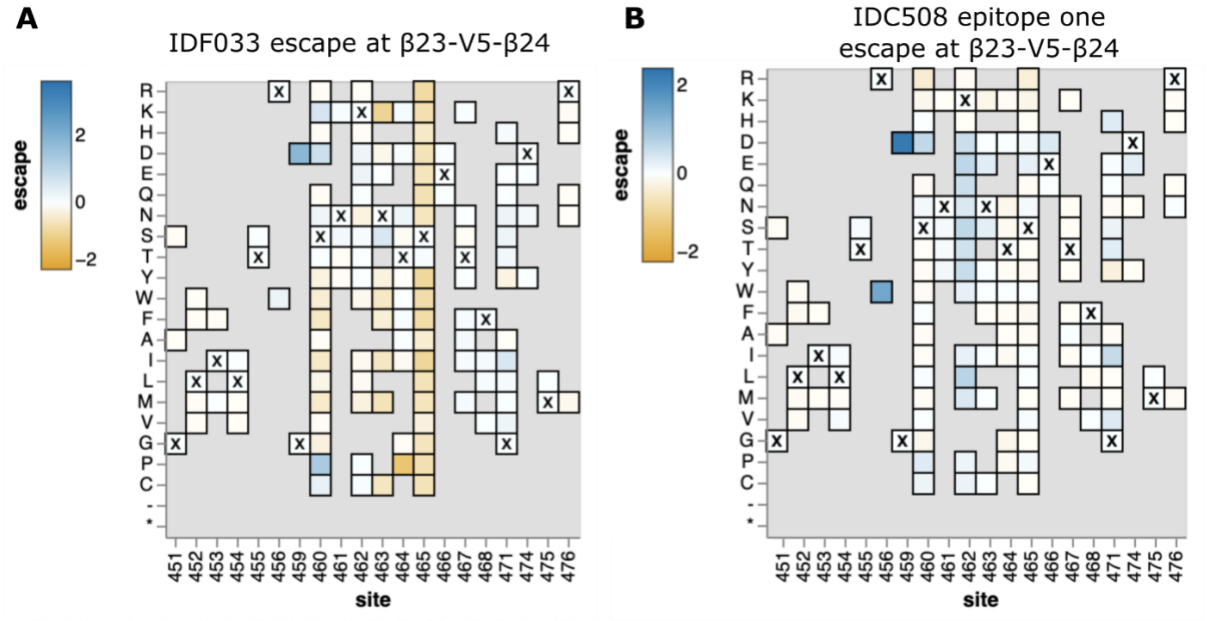


Figure 3.9: (A) Heatmap of escape of individual mutations in β23-V5-β24 for IDF033. Residues marked with X are wildtype residues in BF520. Residues grayed out are not present in the variant libraries. (B) Heatmap for IDC508 epitope 1. See https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/IDF033_escape_plot.html and https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/IDC508_escape_plot.html for interactive versions of the full Env escape maps for IDF033 and IDC508, respectively.

3.3.8 Deep mutational scanning escape maps validate in neutralization assays

We validated the deep mutational scanning by performing pseudovirus neutralization assays on single amino-acid mutants of Env with a range of effects in the escape maps. The changes in pseudovirus neutralization assay IC80's correlated well with the mutational effects predicted by the deep mutational scanning for all four sera (Figure 3.10A,B).

The correlation between the deep mutational scanning and neutralization assays was particularly good for strong escape mutations. For every serum, the tested mutations predicted to most strongly escape neutralization by the deep mutational scanning indeed increased the neutralization assays IC80 (Figure 3.10A,B). The correlation was less consistent for mutations that enhanced neutralization sensitivity rather than escape. For instance, N276D for serum IDC561 caused greater enhancement of neutralization sensitivity in the neutralization assays than predicted from the deep mutational scanning (Figure 3.10A,B). The reduced accuracy of the deep mutational scanning for identifying sensitizing mutations is probably because the mapping experiments were performed at relatively higher serum concentrations (typically exceeding the IC90 for unmutated BF520), making them better suited to identify escape rather than sensitizing mutations.

We also tested Env variants with combinations of mutations in neutralization assays, and again found a good correlation between the deep mutational scanning and neutralization assays (Figure 3.10C,D). For instance, the deep mutational scanning mapped IDC508 to have two epitopes, and pseudovirus neutralization assays using combinations of mutations supports this prediction (Figure 3.10C,D). Specifically, T198D is in one epitope of IDC508 while N276D and G459D are in the other epitope (Figure 3.8A)---and as predicted, N276D and G459D each cause more escape when combined with T198D (Figure 3.10C,D). Note, however, that the effects of these combinations of mutations is complex because N276D has some sensitizing effect on its own (Figure 3.10C,D). It could be that N276D simultaneously causes sensitization to some antibodies but escapes from others, which has been observed for other broad HIV neutralizing serum[142]. The greatest escape from IDC508 is caused by combining all three of T198D, N276D, and G459D, suggesting that this combination escapes a substantial fraction of the neutralizing antibodies in the serum (Figure 3.10C,D).

Figure 3.10 Pseudovirus neutralization assays to validate deep mutational scanning measurements.

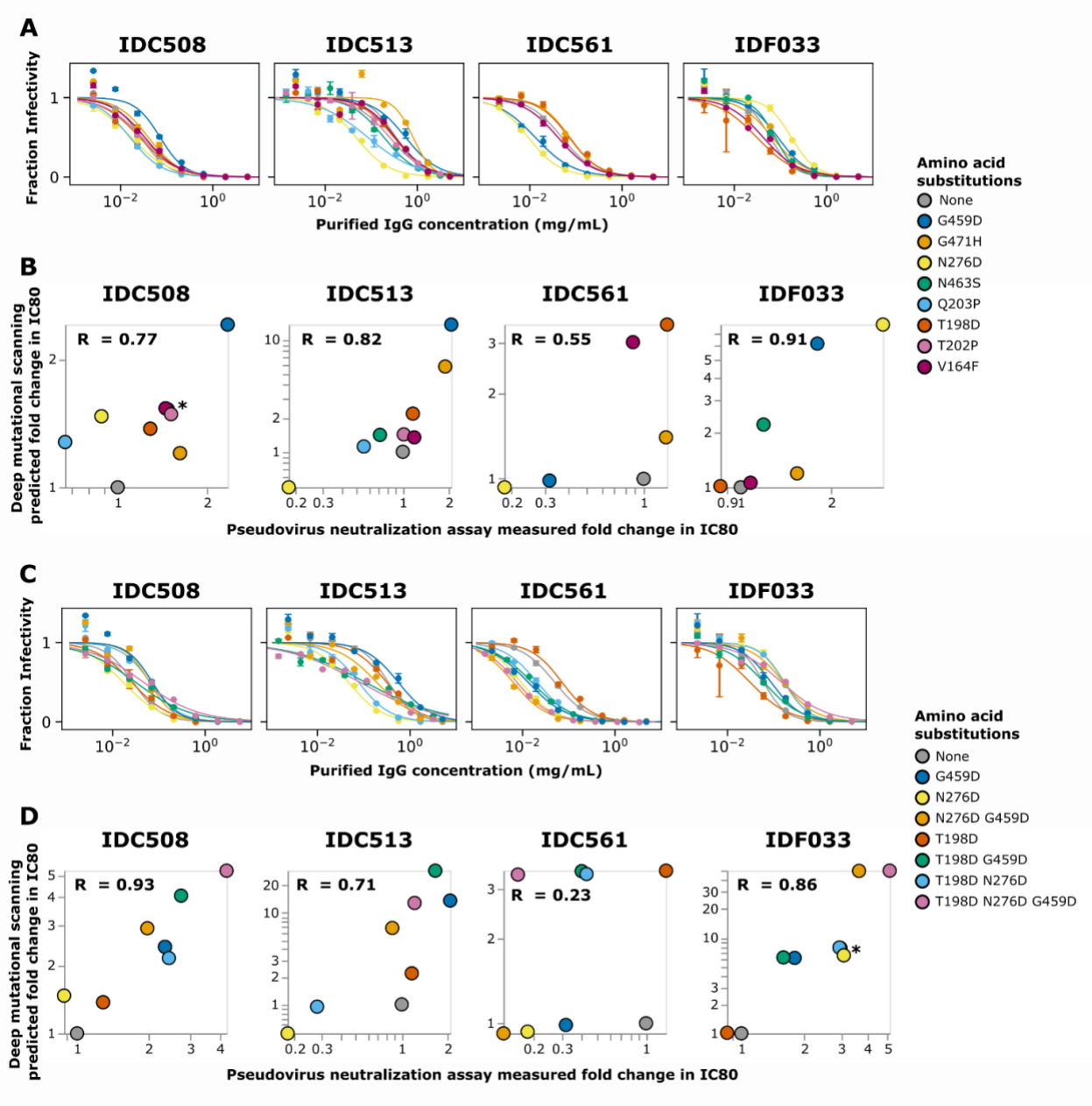


Figure 3.10: (A) Neutralization curves for unmutated BF520 Env and single mutants against purified IgGs from each sera. (B) Correlation of deep mutational scanning predicted fold change IC80s for single mutants versus fold change in IC80 measured in the neutralization assay. R indicates the Pearson correlation. (C) Neutralization curves for combinations of mutations (and the constituent individual mutations repeated from panel A). (D) Correlation of deep mutational scanning predicted fold change IC80s versus fold change IC80s measured in neutralization assays. Points in the scatter plots with asterisks overlap, so have been slightly jittered for clarity.

For sera IDC513 and IDC561, the deep mutational scanning predicted that combinations of mutations would not have substantially more escape than the single mutations with the highest effect, and this validated in neutralization assays (Figure 3.10C,D). Deep mutational scanning mapped IDC513 and IDC561 to each have one epitope (Figure 3.6A, Figure 3.7A). As expected for sera that target a single epitope, no combinations of mutations caused higher fold change IC80 in neutralization assays than the best escaping single mutations for IDC513 and IDC561 (Figure 3.10C,D). For IDC561, only one mutation tested in combinations (T198D) was measured in the deep mutational scanning to be an escape mutation, with the others being sensitizing mutations. Consistent with the deep mutational scanning, only the T198D single mutant caused escape in neutralization assays (Figure 3.10C,D), although similarly to the single mutants the sensitizing effects of N276D and G459D were underestimated by the deep mutational scanning. This poor estimation of the effects of sensitizing mutations leads to a worse overall correlation between the deep mutational scanning and validations for IDC561 (Figure 3.10C,D).

Despite being a sensitizing mutation for some sera, deep mutational scanning predicted mutations to site 276 to cause strong escape from serum IDF033 (Figure 3.8A), and as expected N276D caused a large increase in neutralization assay IC80 both alone (Figure 3.10A,B) and in combination with other mutations (Figure 3.10C,D). Consistent with the deep mutational scanning, combining N276D with another strong escape mutation, G459D, further increased the neutralization assay IC80 (Figure 3.10C,D).

3.4 Discussion:

We have described a lentiviral system for deep mutational scanning measurements of how mutations to Env affect antibody and serum neutralization. A major strength of this system is that it can measure the effects of combinations of mutations, enabling more effective mapping of escape from polyclonal serum that may target multiple epitopes. In addition, since the system

is based on lentiviral vectors that can only undergo a single round of cellular infection, the experiments can be safely performed at biosafety level 2. Therefore, compared to prior Env deep mutational scanning that used fully replicative HIV[47–50], this system both enables more accurate measurements of how Env mutations affect neutralization and is safer and more convenient.

We have used the system to map the neutralization specificities to broadly neutralizing human anti-HIV sera. The mapping shows that although these sera all target Env's CD4 binding site, they differ markedly in the actual epitopes that are the focus of the neutralizing response. Several of the sera have neutralization escape maps that strongly resemble those of characterized monoclonal antibodies. In one case, a serum resembles an antibody (1-18) actually isolated from the same individual. In another case, a serum resembles an antibody (3BNC117) from a different individual, suggesting convergent evolution of the same neutralizing antibody specificity across multiple individuals. However, one serum has an escape map best explained by neutralizing activity contributed by antibodies targeting two distinct epitopes near the CD4 binding site. Key predictions from all of these escape maps validated well in pseudovirus neutralization assays.

Neutralization by each of the sera we mapped is critically affected by the N276 glycan. However, the nature of this dependence is variable: for some sera neutralization is enhanced by disrupting the N276 glycan, where for others neutralization is escaped by disrupting this glycan. These findings are consistent with prior work. For instance, several described cases of broadly neutralizing antibody development have involved gain and loss of the N276 glycan during antibody-virus coevolution[142,173]. Significant progress has been made in germline-targeting vaccinations that attempt to stimulate broadly neutralizing antibody precursors, but such strategies have not yet elicited responses in humans that neutralize heterologous viruses bearing the N276 glycan[40,174,41]. Here we observe strong effects of mutations at this glycan for several broad and potent sera targeting the CD4 binding site, suggesting the presentation of this

glycan may be critical to any vaccination strategies that attempt to elicit such a response. In the future this method can be extended to identify mutations with sensitizing effects similar to N276 glycan knockout mutations, which could further aid sequential vaccine design efforts.

More broadly, the method described here can be used to map polyclonal neutralization escape in a variety of contexts. Combinations of broadly neutralizing antibodies targeting different regions of Env are being explored as potential therapies, and maps of escape from neutralization by single and combinations of antibodies could aid in antibody selection. Vaccine elicited sera can also be mapped to evaluate experimental vaccines and compare their neutralization activity to known broadly neutralizing antibodies or sera. Direct mapping of neutralization escape by polyclonal serum is therefore a useful tool for informing the design of both therapeutics and vaccines.

3.5 Limitations of this study

Our experiments examined the effects of mutations in just one viral Env, BF520. However, mutations can have different effects across strains of HIV[48], so these BF520 escape maps might not fully recapitulate escape mutations in other viral strains. For instance, BF520 is relatively resistant to neutralization by V3-specific antibodies[141], which limits the ability to map escape from such antibodies using the current library. Future studies could extend the work we described here to Envs from more HIV strains to study these differences. Our study was also limited to a moderate number of sera capable of potently neutralizing BF520 Env. In addition, the method described here measures neutralization activity of sera and not other activities such as antibody binding, but it can be used in combination with other techniques like emPEM to study polyclonal serum antibody binding[51].

3.6 Methods

Table 3.1 Primers and sequences

Primers for linearizing BF520 sequences	
Forward linearizing primer (VEP_amp_for_long)	AcgcgCAGCCGAGCCACATCGCTC
Reverse linearizing primer (lin_rev_BF520)	GGAAAGAGTGTgatccaactaggcgc
Primers for mutagenizing BF520 sequences	
Forward mutagenic primer pool	https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/library_design/results/primers/IDT_BF520_primers_for.csv
Reverse mutagenic primer pool	https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/library_design/results/primers/IDT_BF520_primers_rev.csv
Primers for barcoding BF520 sequences	
Forward linearizing primer (VEP_amp_for_long)	AcgcgCAGCCGAGCCACATCGCTC
Reverse barcoding primer (BC_BF520_long)	tatgtaacgcggaactccactaggaacatttctctctcgaaTCTAGANNNNNNNNNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTgatccaactaggcgc
PacBio nucleotide tagging primer set one	
5’ nucleotide tagging primer (PacBio_5pri_G)	ctagccattcagagGCAGCCGAGCCACgTCGCTC
3’ nucleotide tagging primer (PacBio_3pri_C)	CGCTCAACCAGTACGAGCCGTAAGTTATGTAACGCGGAACTCCACcAGGAAC
PacBio nucleotide tagging primer set two	
5’ nucleotide tagging primer (PacBio_5pri_C)	ctagccattcagagGCAGCCGAGCCACcTCGCTC

3’ nucleotide tagging primer (PacBio_3pri_G)	CGCTCAACCAGTACGAGCCGTAAGTTATG TAACGCGGAACTCCACgAGGAAC
PacBio round two primers	
5’ PacBio round 2 forward primer (PacBio_5pri_RND2)	CTAGCCATTCAGAGGCAGCCGAG
3’ PacBio round 2 forward primer (PacBio_3pri_RND2)	CGCTCAACCAGTACGAGCCGTAAGTTATG TAAC
Round one Illumina barcode sequencing preparation primers	
Illumina round 1 forward primer (IlluminaRnd1_For)	CTCTTTCCCTACACGACGCTCTTCCGATC T
Illumina round 1 reverse primer (IlluminaRnd1_rev3)	CTGGAGTTCAGACGTGTGCTCTTCCGATC Tgtccctattggcgttactatgggaacatacgtc
Round two Illumina barcode sequencing preparation primers	
Illumina round 2 universal forward primer (Rnd2ForUniversal)	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACGCTCTTCCGATCT
Illumina round 2 indexing reverse primer (Indexing primers)	CAAGCAGAAGACGGCATACGAGATxxxxxx xxGTGACTGGAGTTCAGACGTGTGCTCTT CCGATCT Where “xxxxxxxx” are NextFlex indices unique to each primer

3.6.1 Design of lentivirus vector backbone

The lentivirus backbone we used is described in Dadonaite et al.[150] See https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/plasmid_maps/lentivirus_backbone_plasmids/pH2rU3_ForInd_mCherry_CMV_ZsGT2APurR.gb for a map of the plasmid containing this backbone. Briefly, the backbone has a repaired 3’ LTR which allows it to be re-rescued after integrating into cells[150], constitutive expression of ZsGreen and puromycin resistance as selectable markers for infection, and a TRE3G promoter that inducibly expresses HIV Env when

the reverse tetracycline transactivator (rtTA) in the 293T-rtTA cells (described in Dadonaite et al.[150] where they are referred to as HEK-293T-rtTAs) is induced by the presence of doxycycline. We used a codon optimized sequence of the HIV Env strain BF520.W14M.C2[151,152]. See https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/plasmid_maps/lentivirus_backbone_plasmids/pH2rU3_ForInd_BF520.gb for a plasmid map containing the codon optimized BF520 sequence.

3.6.2 Design of mutant libraries containing mostly functional mutants

To choose mutations to include in our mutant libraries based on prior BF520 deep mutational scanning[48], we compared previously measured effects of all mutations vs the effects of stop codons. We retained mutations with an effect measured to be more positive than the 0.95 quantile of stop codon effect in the previous deep mutational scanning. We only retained three stop codons, at sites 100, 200, and 300, so we could use these as controls for selections. We wanted to include mutations present in natural HIV sequences even if they had negative effects in previous deep mutational scanning, since these mutations are tolerable when combined with some other mutations and we wanted our neutralization selections to include most naturally occurring mutations. We downloaded the 2018 filtered web alignment of group M HIV-1 sequences without recombinants from the Los Alamos HIV sequence database[160], and used it to identify any mutations relative to BF520 that were present more than once in the alignment, and retained these mutations for the mutant libraries in addition to those chosen above.

See https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/tree/main/library_design for the analysis to choose these mutations. See https://github.com/dms-

vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/library_design/results/IDT_library_df.csv for the retained mutations.

3.6.3 Design of primers for BF520 mutagenesis

See https://github.com/jbloomlab/TargetedTilingPrimers for the script used to generate primer sequences to make the chosen mutations. This script generates forward and reverse primers for each mutation which mutate that site to the most frequent codon of the desired mutant. Primer pools were ordered as oPools from Integrated DNA Technologies. See https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/tree/main/library_design/results/primers for the primer sequences.

3.6.4 Production of plasmids containing barcoded mutant BF520 sequences

See https://github.com/jbloomlab/CodonTilingPrimers for a general description of the PCR mutagenesis strategy we use here. The key difference is that we only ordered primers that introduced the targeted amino-acid mutations.

To mutagenize the BF520 sequences, a codon optimized BF520 sequence was first amplified from a plasmid containing the codon optimized BF520 sequence in a lentiviral backbone. See https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/plasmid_maps/lentivirus_backbone_plasmids/pH2rU3_ForInd_BF520.gb for the sequence of this plasmid. The PCR was performed with the following conditions: PCR mix: 18.5 µL H2O, 2.5 µL DMSO (to reduce off-target amplification), 1.5 µL 10µM forward linearizing primer (VEP_amp_for_long) , 1.5 µL 10 µM reverse linearizing primer (lin_rev_BF520), 1 µL 10ng/µL BF520 template plasmid, and 25 µL 2x KOD Hot Start Master Mix (Sigma-Aldrich, Cat. No. 71842). Cycling conditions: (1) 95C/2min (2)

95C/20sec (3) 70C/1sec (4) 54C/10sec, cooling at 0.5C/sec (5) 70C/40sec (6) Return to Step 2 x19.

The amplified, linearized BF520 sequence was gel purified using NucleoSpin Gel and PCR Clean-up kit (Takara, Cat. No. 740609.5) and then purified using Ampure XP beads (Beckman Coulter, Cat. No. A63881) at 1:1 sample to bead ratio.

The amplified BF520 sequence was then used in a modification of a previously described PCR mutagenesis technique[45]. Forward and reverse pools of codon tiling primers for generating specific mutations were generated using https://github.com/jbloomlab/TargetedTilingPrimers, as described above. In separate PCR reactions, the forward primer pool was used with the reverse linearizing primer and the reverse primer pool was used with the forward linearizing primer. The conditions for these PCR reactions were as follows: PCR mix: 7.7 µL H2O, 1.5 µL DMSO, 4 µL 3 ng/µL linearized BF520 template, 0.9 µL 10 µM forward or reverse primer pool, 0.9 µL reverse (lin_rev_BF520) or forward (VEP_amp_for_long) linearizing primer, and 15 µL 2x KOD Hot Start Master Mix. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 50C/20sec, cooling at 0.5C/sec (50 70C/120sec (6) Return to Step 2 x9.

After the mutagenic PCRs, a joining PCR was performed using products from the forward and reverse primer pool mutagenic PCRs. The conditions for the joining PCRs were as follows: PCR mix: 4µL H2O, 4µL forward primer pool mutagenesis PCR product diluted 1:4 with H2O, 4µL reverse primer pool mutagenesis PCR product diluted 1:4 with H2O, 1.5µL 5µM forward linearizing primer (VEP_amp_for_long) , 1.5µL 5µM reverse linearizing primer (lin_rev_BF520) , and 15µL 2x KOD Hot Start Master Mix. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 50C/20sec, cooling at 0.5C/sec (5) 70C/120sec (6) Return to Step 2 x19.

The resulting mutagenized BF520 sequences were gel purified and Ampure bead cleaned with a 1:1 product to beads ratio. These mutagenized sequences were then barcoded

with random nucleotide barcodes using a PCR with the following conditions: PCR mix: 30 ng joining PCR product, 1.5µL 5µM forward linearizing primer (VEP_amp_for_long), 1.5µL 5µM reverse barcoding primer (BC_BF520_long), 15 µL 2x KOD Hot Start Master Mix, and fill to 30 µL with H2O. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 50C/10sec, cooling at 0.5/sec (5) 70C/120sec (6) Return to Step 2 x9.

The barcoded mutagenized BF520 sequences were gel and Ampure bead purified, and then cloned into a lentiviral backbone containing plasmid as described in Dadonaite et al.[150], with some modifications as follows. The barcoded mutagenized sequences were first cloned into an earlier version of the lentiviral backbone during system development. The map of the plasmid used can be found at https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/plasmid_maps/lentivirus_backbone_plasmids/pH2rU3_ForInd_mCherry_CMV_ZsG_NoBC_cloningvector.gb. The plasmid was digested with MluI and XbaI, and then gel and Ampure bead purified. The barcoded mutagenized BF520 sequences and the digested plasmid were eluted into H2O after Ampure bead purification, which we have observed results in higher Hifi assembly efficiency. We then used a 2:1 insert to vector ratio in a 1 hour Hifi assembly reaction using NEBuilder HiFi DNA Assembly kit (NEB, Cat. No. E2621). The Hifi assembly products were Ampure bead purified and eluted into 20 µL of H2O, which we have observed results in a higher electroporation efficiency. We used 2 µl of the purified HiFi product to transform 20 µl of 10-beta electrocompetent E. coli cells (NEB, C3020K). We performed 5 electroporation reactions for a final count of >5 million CFUs per library. We aimed for this high diversity of barcoded mutants in transformants to reduce the potential of barcode sharing in virus libraries, which we will describe in detail below. We plated the transformed cells on LB+ampicillin plates, incubated the plates overnight at 37 C, and scraped the plates the next day to collect the transformants. The OD600 of the collected bacteria were measured, and the bacteria were diluted to 15 OD600 and used in five separate 5 mL minipreps (QIAprep Spin Miniprep Kit, Cat. No. 27106X4) each,

resulting in a total of ~200 ug of plasmid being isolated for each replicate library. The rest of the bacteria were spun down in pellets and stored.

At a later stage of system development, we decided to move the barcoded mutagenized sequences into an improved version of the lentiviral backbone that uses puromycin selection rather than flow cytometry sorting to enrich infected cells when making the integrated mutant library cell lines. The map of this plasmid can be found at https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/plasmid_maps/lentivirus_backbone_plasmids/pH2rU3_ForInd_mCherry_CMV_ZsGT2APurR.gb. We decided to do this using restriction digest and ligation cloning of the library plasmids and the new lentiviral backbone plasmid. As an important note for future deep mutational scanning studies, this cloning strategy was not optimal. Since the barcoded mutagenized sequences were drawn from a plasmid pool with relatively limited diversity compared to mutagenic PCR products (a few million unique barcoded sequences vs >>billions of unique barcoded sequences), this cloning imposed an additional unintended bottleneck on the barcoded mutagenized sequence diversity. This meant that the final plasmid pools for each library had lower barcode diversity than intended, resulting in some degree of barcode sharing, described in a lower section. In the future, it is advised for similar deep mutational scanning strategies aiming for extremely high plasmid diversity to only clone from highly diverse mutagenic PCR products rather than any pre-existing mutant plasmid pool, which will always be limited in diversity by transformation efficiencies.

To move the barcoded mutagenized sequences into the improved lentiviral backbone, we digested each mutant plasmid pool and the new lentiviral backbone using MluI and XbaI. We gel extracted and Ampure bead cleaned the mutagenized barcoded inserts from the mutant plasmid pools and the cut lentiviral backbone vector, and eluted in Qiagen EB buffer (Cat. No. 19086). We then used T4 DNA ligase (New England BioLabs, Cat. No. M0202S) to ligate the inserts with the vector, using the following conditions: Reaction mix: 2 µL T4 DNA Ligase Buffer (10x), 50 ng Vector DNA, 45.35 ng insert DNA, 1 µL T4 DNA Ligase, and fill with H2O to 20 µL.

We incubated the reaction at room temperature for 10 minutes, heat inactivated at 65C for 10 minutes, and then Ampure bead cleaned the product and eluted in 20 µL H2O. We then electroporated NEB 10beta cells (New England BioLabs, Cat. No. C3020K) following the protocol (https://www.neb.com/protocols/0001/01/01/electroporation-protocol-c3020) exactly. We performed five electroporations per library, for a total of ~1 million CFUs per library. Again, as a note to future deep mutational scanning studies, the mutant plasmid pool restriction digest and ligation cloning strategy used here along with a transformation bottleneck <5 million CFUs is not recommended due to potential unintended bottlenecking of barcoded mutants.

3.6.5 Production of cell lines storing BF520 mutant libraries

Production of cell line-stored BF520 mutant libraries was performed similarly to previously described in Dadonaite et al.[150], with modifications (Figure 3.1B). This process involved the same steps of: 1) production of VSV-G pseudotyped lentiviruses carrying the barcoded mutant BF520 sequences, 2) infection of 293T-rtTA cells with the VSV-G pseudotyped viruses, and 3) selection for transduced cells using puromycin.

In order to not bottleneck the diversity of barcoded mutants at this step, we aimed to produce many more VSV-G pseudotyped viruses carrying the barcoded mutant BF520 sequences than the eventual desired library sizes of around 40,000 barcoded variants. We plated 500,000 293T cells per well in 6 well plates, and transfected 12 wells for each library. We used BioT (Bioland Scientific) for the transfections, and followed the manufacturer's recommendations for the protocol and DNA / transfection reagent ratios. We transfected each well with 1 ug of lentiviral backbone plasmids carrying the barcoded mutagenized BF520 sequences, 250 ng of a HIV Tat expressing plasmid (HDM-tat1b), 250 ng of a HIV Rev expressing plasmid (pRC-CMV_Rev1b), 250 ng of a HIV Gag-Pol expressing plasmid (HDM-Hgpm2), and 250 ng of a VSV-G expressing plasmid (HDM_VSV_G). See https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/tree/main/plasmid_maps

for maps of these plasmids. We pooled the transfection supernatents for each library 48 hours post-transfection, filtered through a 0.45 µm SFCA syringe filter (Corning, Cat. No. 431220), and stored in 1 mL aliquots at -80C. We titrated these viruses based on the percent ZsGreen expression of cells infected with dilutions of virus as determined by flow cytometry, as described in Crawford et al.[70] This yielded a total of >20 million viruses per library.

We used these VSV-G pseudotyped viruses to infect 293T-rtTA cells with approximately the same number of viruses as barcoded mutants that we desired in the final virus libraries. We aimed to avoid any bottlenecks in the barcoded mutant sequences before this step because recombination of pseudodiploid lentiviral genomes and mutations caused by lentiviral reverse transcription will alter barcode-mutant linkage during this step[71–73,150]. We attempted to maintain high diversity in the barcoded sequences in prior steps to ensure each barcoded mutant-carrying lentiviral genome would have a unique barcode, so that barcodes would not be repeated in infected cells. After this step, each cell in the library storing cell lines will only have one integrated lentiviral genome with one barcoded mutant, so recombination in future steps is not an issue and mutations caused by reverse transcription in future steps will not alter mutant BF520 expression from these integrants and can be filtered in PacBio sequencing data, described below.

We aimed to infect the 293T-rtTA cells with between 30,000-40,000 variants per library. We first plated 500,000 293T-rtTA per well in ten six well plates. The next day, at the time of infection, we counted the cells per well in several wells. Based on the average count, we infected each well with the amount of infectious units required for a 0.005 multiplicity of infection, for five six well plates per library. Two days later, we determined the actual multiplicity of infection and infectious units per well for each library by determining the percent of infected cells by flow cytometry on ZsGreen expression and back-calculating the infectious units added per well based on that percentage and the average cell count per well at the time of infection.

For each library we then pooled cells from the number of wells required for a total infectious units between 30,000-40,000. The pooled cells for each library were plated in a 10 cm plate.

Transduced cells were then selected for using puromycin selection, since infected cells expressed the puromycin resistance gene from the lentiviral genome while non-infected cells did not. Puromycin was added 24 hours after pooling at 0.75 ug/mL. 48 hours later, the cells were split into three 15 cm dishes per library with 0.75 ug/mL puromycin. 48 hours later, the media was replaced with fresh media plus 0.75 ug/mL puromycin. 48 hours later (a week after pooling), the cells for each library appeared all ZsGreen positive under a fluorescent microscope, and were expanded into one five layer flask (Falcon, Cat. No. 353144) per library. 24 hours later, half of the cells per library were frozen in 1 mL aliquots of 5 million cells in tetracycline-negative heat-inactivated fetal bovine serum (Gemini Bio, Cat. No. 100-800) with 10% DMSO, to be used in future virus library generation. The rest of the cells were used to generate mutant virus libraries as described below.

3.6.6 Production of BF520 and VSV-G pseudotyped mutant virus libraries

Since each cell in the cell lines produced as described above contained one barcoded BF520 mutant, we were able to produce genotype-phenotype linked BF520 mutant virus libraries from them (Figure 1B). We did this by plating 100 million cells per flask in two five layer flasks per library in 150 mL of tetracycline free D10. 24 hours later, we transfected each flask using BioT by using 225 µL of BioT mixed with 7.5 mL of DMEM and a DNA mix containing 50 ug of each lentivirus helper plasmid (Tat, Rev, and Gag-Pol). We also induced Env expression at the time of transfection by adding doxycycline to a final concentration of 100 ng/mL. 48 hours later, the supernatant for each library was filtered through a 0.45 µM SFCA filter (Nalgene, Cat. No. 09-740-44B). The filtered virus was then concentrated using ultracentrifugation with a 20% sucrose cushion at 100,000 g for one hour. The viruses were resuspended in 500 µL of DMEM,

and were typically around ten million infectious units per mL. We then stored these viruses at -80C.

We also generated VSV-G pseudotyped viruses from the library cell lines to use for PacBio sequencing and as controls for selections on the effects of mutations on BF520 function, described below. We did this by plating four million cells per plate in three 10 cm dishes for each library, and transfecting each plate 24 hours later using BioT according to the manufacturer's recommendations. For the DNA mix, we used 2.5 ug of each lentivirus helper plasmid (Tat, Rev, and Gag-Pol) and a VSV-G expressing plasmid (four plasmids, 10 ug total DNA) per plate. 48 hours later we pooled the supernatants for each library and filtered them through a 0.45 µM SFCA filter. We then stored these viruses at -80C.

3.6.7 PacBio sequencing of mutants present in mutant libraries

We used long-read sequencing PacBio sequencing to simultaneously determine the composition of the mutant libraries contained in the library cell lines and link mutants with their random nucleotide barcodes. First, we plated 1 million 293Ts per well in poly-L-lysine coated six well plates (Corning, Cat. No. 356515). 24 hours later, we infected two wells of cells with 1 million infectious units of +VSV-G library virus per well, for each library. Six hours later, we removed the media, washed the cells with PBS, and miniprepped each well, which isolates unintegrated lentivirus genomes as described previously.[47,150] Each well was miniprepped independently and eluted using 50 µL of EB.

A two-step PCR strategy was then used to amplify the barcoded mutant BF520 sequences for PacBio sequencing, as described previously.[150] Briefly, the miniprepped products for each library were split into two short-cycle initial PCRs that attached single nucleotide tags to each end of the amplicon that were unique for each PCR. The products of these initial PCRs were then pooled for each library for longer cycle PCRs to amplify enough DNA for PacBio sequencing. The single nucleotide tags from the initial PCRs then allowed us to later estimate

the amount of strand exchange that occurred in the longer cycle PCRs based on the frequency of tags found together in PacBio sequences that were from different first round PCRs. The first round PCR is a low cycle number to minimize the probability of strand exchange during it, and the number of cycles in the second PCR was lowered as much as possible to minimize strand exchange while still generating enough DNA for PacBio sequencing. Here are the conditions used for the first round of PCRs: PCR mix: 10 µL of miniprep product, 1 µL of 10 µM 5' nucleotide tagging primer (PacBio_5pri_G or PacBio_5pri_C), 1 µL of 10 µM 3' nucleotide tagging primer (PacBio_3pri_C or PacBio_3pri_G), 20 µL KOD Hot Start Master Mix, and 8 µL H2O. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 60C/10sec, cooling at 0.5/sec (5) 70C/60sec (6) Return to Step 2 x7 (7) 70C/60sec.

The PCR products were cleaned with Ampure beads with a 1:1 product to beads ratio and eluted into 35 µL of EB. We then used the following conditions for the second round of PCRs: PCR mix: 10.5 µL of first variant tag set round 1 PCR product, 10.5 µL of second variant tag set round 1 PCR product , 1 µL of 10 µM 5' PacBio round 2 forward primer (PacBio_5pri_RND2), 1 µL of 10 µM 3' PacBio round 2 reverse primer (PacBio_3pri_RND2), and 25 µL KOD Hot Start Master Mix. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 60C/10sec, cooling at 0.5/sec (5) 70C/60sec (6) Return to Step 2 x10 (7) 70C/60sec. The PCR products were Ampure bead cleaned, and each eluted into 40 µL of EB. The cleaned products for each library were pooled. Each library pool was then barcoded for PacBio sequencing using SMRTbell prep kit 3.0, bound to polymerase using Sequel II Binding Kit 3.2, and then sequenced using a PacBio Sequel IIe sequencer with a 20-hour movie collection time. The data were analyzed as described below (section "PacBio sequencing data analysis").

3.6.8 Barcode amplification for Illumina sequencing of mutants after selections

After the above step using PacBio sequencing to link each mutant and barcode, future experimental steps only require short read sequencing of barcodes to determine changes in variant frequencies across conditions. We amplified barcodes for sequencing as previously described in Dadonaite et al.[150] with slight modifications, repeated here. A first round of PCR was used to amplify the barcodes using a forward primer that aligns to the Illumina Truseq Read 1 sequence upstream of the barcode in our lentiviral backbone and a reverse primer that annealed downstream of the barcode and overlapped with the Illumina Truseq Read 2 sequence. This PCR used the following conditions: PCR mix: 22 µL of miniprepped selection sample, 1.5 µL of 10 µM 5' Illumina round 1 forward primer (IlluminaRnd1_For), 1.5 µL of 10 µM 3' Illumina round 1 reverse primer (IlluminaRnd1_rev3), and 25 µL KOD Hot Start Master Mix. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 58C/10sec, cooling at 0.5/sec (5) 70C/20sec (6) Return to Step 2 x27.

The PCR products were Ampure bead cleaned with a 1:3 product to beads ratio, and then DNA concentration was quantified using a Qubit Fluorometer (ThermoFisher). A second round of PCR was then performed using a forward primer that annealed to the Illumina Truseq Read 1 sequence and had a P5 Illumina adapter overhang, and reverse primers from the PerkinElmer NextFlex DNA Barcode adaptor set that annealed to the Truseq Read 2 site and had the P7 Illumina adapter and i7 sample index. This PCR used the following conditions: PCR mix: 20 ng of round 1 product as determined by Qubit, 2 µL of 10 µM 5' Illumina round 2 universal forward primer (Rnd2ForUniversal), 2 µL of 10 µM 3' Illumina round 2 indexing reverse primer (Indexing primers), 25 µL KOD Hot Start Master Mix, and fill to 50 µL total using H2O. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 58C/10sec, cooling at 0.5/sec (5) 70C/20sec (6) Return to Step 2 x19.

The DNA concentration of each round 2 PCR product was quantified using Qubit. The samples were pooled at an even ratio, gel purified and Ampure bead cleaned at a 1:3 sample to beads ratio, and then sequenced using either P2 or P3 reagent kits on a NextSeq 2000. The data were analyzed as described below (section “Illumina barcode sequencing data analysis”).

3.6.9 Selections on effects of mutations on the function of BF520

To measure the effects of mutation on BF520 mediated entry into cells, we infected cells with VSV-G and non-VSV-G pseudotyped mutant virus libraries separately. To do this, we plated 1 million TZM-bl cells in each well of six well plates. 24 hours later, we infected each well with ~1 million infectious units of VSV-G or non-VSV-G pseudotyped mutant virus depending on the condition. We used this amount of virus because we aimed to use >20x the size of each mutant library during infections, so that each barcoded mutant would be present more than once and less likely to be randomly bottlenecked during the selections. During infections, we added 100 ug/mL DEAE dextran, which improves the infectivity of Env pseudotyped viruses and results in less random bottlenecking of mutants during infections.[47,49] 12 hours after infection, the cells were washed with PBS, miniprepped using a QIAprep Spin Miniprep Kit to isolate unintegrated lentivirus genomes as described previously[47,150], and eluted into 30 µL of EB. To improve the DNA recovery, the EB was run through the column twice, incubating at 55C for five minutes before spinning each time. The eluent was then used in the barcode sequencing prep described above.

3.6.10 Production of VSV-G pseudotyped standard viruses for neutralization selections

For each selection using antibodies or sera, we spiked in a small amount of a separately produced only-VSV-G pseudotyped virus pool carrying known barcodes to act as neutralization standards by enabling conversion of barcode counts to absolute neutralization values (See Figure 2.6). We produced these viruses exactly as described in Dadonaite et al.[150] Briefly, 293T-

rtTA cells were transduced at a low multiplicity of infection with a pool of lentiviruses carrying a small set of known barcodes but no viral entry protein in their genomes. Transduced cells were selected for using flow cytometry cell sorting on ZsGreen expression, and then standard viruses were produced by transfecting the cells with the lentiviral helper plasmids and a plasmid expressing VSV-G. The result of this process was a standard virus pool with known barcodes that was produced in the same manner as mutant libraries but did not contain any viral entry protein mutants.

3.6.11 Selections on effects of mutations on neutralization escape

We aimed to perform antibody and serum selections at concentrations between the IC90-IC99.9 for each antibody and serum. We used a spread of concentrations in this range because it is difficult to estimate IC9X concentrations and we wanted to use a spread of high neutralization levels to fit our biophysical escape models.[78] When performing selections using antibodies or serum with the mutant virus libraries, we spiked-in the VSV-G pseudotyped neutralization standard viruses to be 0.5-1% of the total infectious units in the virus pool. From this combined virus pool, 1 million infectious units per selection were incubated with antibody or serum at the desired concentration for one hour. After the incubation, the volume of each condition was raised to 2 mL with 100 ug/mL DEAE dextran using D10 with the appropriate amount of DEAE dextran. Each condition was used to infect one well of TZM-bl cells in a six well dish plated at 1 million cells per well 24 hours prior. 12 hours after infection, the cells were washed with PBS, miniprepped, and eluted into 30 µL of EB. To improve the DNA recovery, the EB was run through the column twice, incubating at 55C for five minutes before spinning each time. The eluent was then used in the barcode sequencing prep described above.

3.6.12 Validation pseudovirus neutralization assays

Plasmids containing BF520 with mutations used in pseudovirus neutralization assays were ordered from Twist in the HDM plasmid (https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/plasmid_maps/viral_entry_protein_expression_plasmids/HDM_BF520.gb). To produce viruses pseudotyped with each BF520 mutant, we first plated 500,000 293T cells per well in six well plates. 24 hours later we transfected 1 ug of a ZsGreen and Luciferase expressing lentivirus backbone plasmid (https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/plasmid_maps/lentivirus_backbone_plasmids/pHAGE6-wtCMV-Luc2-BrCr1-ZsGreen-W-1247.gb), 250 ng of each lentiviral helper plasmid (Tat, Rev, and Gagpol), and 250 ng of the HDM plasmid expressing the desired BF520 mutant into each well. We collected the viruses 48 hours later by filtering the supernatant through a 0.45 µm SFCA syringe filter and storing the virus at -80C.

To titrate these viruses for use in neutralization assays, we first plated 25,000 TZM-bl cells per well in clear bottom, poly-L-lysine coated, black walled 96 well plates (Greiner, Cat. No. 655930). 24 hours later, we serially diluted each mutant BF520 pseudotyped virus and infected the cells. 48 hours after infection, we used the Bright-Glo Luciferase Assay System (Promega, E2610) to measure relative light units (RLUs) for each dilution. We estimated the average RLU/µL for each BF520 mutant within a linear range based on its dilution curve. Note, this method and the following described neutralization assay are not the same as a typical TZM-bl neutralization assay, since Luciferase expression will be driven from the lentiviral genome of the infecting virus rather than the pre-integrated Tat-driven Luciferase in the TZM-bl cells, as there is will be no Tat expressed from these lentiviruses. We chose to do this rather than using ΔEnv HIV pseudoviruses in typical TZM-bl neutralization assays so that there was no chance of the BF520 Env mutants with combinations of escape mutations to CD4 binding site antibodies or sera recombining into full-length replicative HIV.

For neutralization assays, we plated 25,000 TZM-bl cells per well in clear bottom, poly-L-lysine coated, black walled 96 well plates. 24 hours later, we serially diluted each antibody or sera, and then incubated each dilution with each mutant BF520 pseudotyped virus for one hour. We then added an equal volume of D10 with DEAE dextran to a final DEAE dextran concentration of 100ug/mL, and infected the TZM-bls. 48 hours later, we used the Bright-Glo Luciferase Assay System to measure RLUs for each dilution.

To calculate fraction infectivity, we subtracted the average background reading of RLUs from uninfected cells from each condition, and then divided the RLU of each antibody or serum dilution by the average RLUs from cells infected by virus that was incubated with media rather than antibodies or sera. The fraction infectivities were used to fit neutralization curves using neutcurve (https://jbloomlab.github.io/neutcurve/). We compared fold change IC80 rather than IC50 for our interpretation of the neutralization assays because our deep mutational scanning selections were performed at high levels of neutralization (>IC90 for wildtype BF520).

3.6.13 Experimental replicates

We performed one or two replicates of each selection with two independent mutant libraries for each experiment. See https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/avg_muteffects.html for correlation plots of functional effects of mutations across replicates and see https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/avg_antibody_escape.html for correlation plots of escape effects of mutations across replicates for each antibody and serum. Throughout the paper we report the median across these replicates.

3.6.14 Cell lines

HEK-293T cells were from ATCC (CRL3216), TZM-bl cells were from the NIH AIDS Reagent Program (ARP-8129, contributed by Dr. John C. Kappes, Dr. Xiaoyun Wu and

Tranzyme Inc.), and 293T-rtTA expressing cells were produced as previously described in Dadonaite et al.[150] where they are referred to as HEK-293T-rtTAs. All cell lines were grown in D10 media (Dulbecco's Modified Eagle Medium with 10% heat-inactivated fetal bovine serum, 2 mM l-glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin). To avoid rtTA activation and mutant BF520 expression earlier than intended, 293T-rtTA cells were grown in D10 made with tetracycline-free fetal bovine serum (Gemini Bio, Cat. No. 100-800).

3.6.15 Antibodies

3BNC117 IgGs were a gift from Dr. Michel Nussenzweig and Dr. Marina Caskey, and PGT151 and 1-18 IgGs were produced by Genscript based on publicly available sequences.

3.6.16 Patient plasma samples and IgG isolation

Blood samples were obtained under protocols approved by the Institutional Review Board (IRB) of the University of Cologne (protocols 13-364 and 16-054) and the local IRBs and all participants provided written informed consent. HIV-1 infected patients were recruited at private practices and/or hospitals in Germany. Plasma samples were obtained and stored at -80°C until further use. Prior to IgG isolation, plasma samples were heat-inactivated at 56°C for 40 minutes. IgGs were isolated through an overnight incubation with Protein G Sepharose (GE Life Sciences) at 4°C, followed by elution with 0.1 M glycine (pH=3.0) using chromatography columns. The eluted IgGs were buffered in 1 M Tris (pH=8.0) and then underwent buffer exchange to phosphate-buffered saline (PBS) and concentration using Amicon 30 kDa spin membranes (Millipore). The purified IgGs were stored at 4°C until further use.

3.7 Computational Analysis

3.7.1 Computational pipeline overview

For analyzing deep mutational scanning of viral entry protein, we use a common, modular pipeline. See https://github.com/dms-vep/dms-vep-pipeline for this pipeline. For this paper, we used version 2.0.1 of dms-vep-pipeline. We created a repository for the analyses performed in this paper. See https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera for this repository. This repository includes the main dms-vep-pipeline as well as all of the scripts, notebooks, and settings necessary to recreate the analysis. Some key results files can be found in this repository, but some results files that are too large are not tracked in the online repository. The pipeline also produces HTML rendering of the key analyses and interactive plots. See https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/ for these pages. These pages are the best way to explore the analyses and interactive plots of the results.

3.7.2 PacBio sequencing data analysis

We used alignparse (see https://jbloomlab.github.io/alignparse/ for documentation) to analyze the PacBio sequencing data[138]. The PacBio CCSs went through several filtering steps before we determined which BF520 mutants were linked to which barcodes. First, we looked for evidence of strand exchange during the PacBio sequencing prep PCRs by computing the fraction of CCSs that contained unexpected pairs of single nucleotide tags, such as pairs of nucleotide tags from different round one PCRs or any wildtype nucleotides. These sequences represented just <1% of the CCSs, and were filtered out.

Next, we computed empirical accuracies for each CCS, which represent the fraction of CCSs with the same barcode that report the same BF520 sequence. The empirical accuracies were around 0.60, slightly less than our previously described SARS-2 spike libraries[150].

Inaccuracies in the BF520 sequences are due to a combination of factors including reverse transcription errors, sequencing errors, strand exchange during PCR, and, importantly, actual linkage of the same barcode sequence with two or more BF520 mutants due to having the same barcode with different BF520 mutants integrated into different cells. As noted in above sections, we know that we unintentionally bottlenecked the barcoded BF520 mutants to a lower level than desired, which likely resulted in some barcode sharing between variants and contributed to the reduction in empirical accuracies.

We filtered out any barcodes that had less than three CCSs or had minor fractions of substitutions or indels above 0.4. This removes consensus sequences that we are not confident in due to not having enough CCSs for that barcode or evidence of multiple BF520 mutants sharing that barcode. The remaining consensus sequences of barcoded mutant BF520 sequences were then saved in barcode / variant lookup table files. See https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/build_pacbio_consensus.html for the analysis described in this section and https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/results/variants/codon_variants.csv for the barcode / variant lookup tables.

3.7.3 Illumina barcode sequencing data analysis

We used the parser found at https://jbloomlab.github.io/dms_variants/dms_variants.illuminabarcodeparser.html/ to determine the counts of each variant in each selection condition. For each experiment, barcoded mutants were only retained if their “pre-selection” counts, such as counts in no-antibody conditions for antibody selections or VSV-G condition counts for functional selections, were above thresholds specified in the configuration file for the analysis. See https://github.com/dms-vep/HIV_Envelope_BF520_DMS_CD4bs_sera/blob/main/config.yaml for the configuration file. This was done because these barcoded mutants are more likely to be randomly bottlenecked

during the infection step of selections and to have highly noisy scores due to the computation of functional and antibody escape scores (described below).

3.7.4 Modeling the effects of mutations on Env function

We modeled the functional effects of mutations on BF520 as described previously in Dadonaite et al.[150] Briefly, we computed functional scores for each barcoded mutant m defined as $log_2\left(\left[n^m{}_{post} \,/\, n^{wt}{}_{post}\right]/\left[n^m{}_{pre} \,/\, n^{wt}{}_{pre}\right]\right)$, where where $n^m{}_{post}$ is the count of mutant m in the post-selection (BF520-pseudotyped) infection, $n^m{}_{pre}$ is the count of mutant m in the pre-selection (VSV-G-pseudotyped) infection, and $n^{wt}{}_{post}$ and $n^{wt}{}_{post}$ are the counts of all unmutated (wildtype) barcoded variants in each condition. We then used global epistasis models[76,77] to deconvolve the functional scores of these thousands of mutants with combinations of mutations into estimates of the effects of each individual mutation. Under these models, each mutation has a "latent effect" on a "latent phenotype" defined as an unmeasured phenotype where mutations interact additively. These latent effects and the latent phenotype score for a mutant are then transformed to the actually measured "observed phenotype" through a nonlinear function. This process results in more accurate estimates for the effects of mutations by modeling some of the epistasis between mutations through the nonlinear relationship between the latent and observed phenotypes. See https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/fit_globalepistasis.html for the models, see https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/muteffects_latent_heatmap.html for the latent effects of mutations averaged across selections, and see https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/muteffects_observed_heatmap.html for the observed effects of mutations averaged across selections. For Figure 3.3C, we used the

observed effects of mutations, since these are the most accurate representation of the effects of mutations on BF520.

3.7.5 Modeling the effects of mutations on antibody and serum escape

We modeled the effects of mutations on antibody and serum escape as described previously in Dadonaite et al.[150] Briefly, we calculated the non-neutralized fraction of each variant at each antibody calculation to get the probability of escape of each variant at each concentration. We then used the software package polyclonal[78] version 4.1 (see https://jbloomlab.github.io/polyclonal/ for documentation) to estimate the effects of each individual mutation on escape using a biophysical model. Under this model, antibodies, mixtures of antibodies, or polyclonal serum can target Env at one or more epitopes. Measured probability of escape of each variant is modeled as a result of how the mutations it has escape the neutralization activity of each epitope. The measured probability of escape of combinations of mutations across mutants is used to optimize the number and sites of antibody epitopes and give each mutation escape scores corresponding to its contributions to escape for each epitope when present. Mutations that have no effect on escape will have scores of zero, while mutations that cause escape will have scores >0. The summed escape scores for each site are the y-axis values displayed in the line plots in each figure and used to color the PDB structures seen in each figure. The individual escape scores for each mutation can be seen in the heatmaps of the linked interactive plots, like the ones seen in Figure 3.4B and Figure 3.9.

The models are also able to predict arbitrary inhibitory concentrations for Env mutants, such as an IC50 or IC80 for serum IDC508 against BF520 with mutations T198D and N276D. This is done by determining the effect of each mutation on escape from each epitope in the neutralizing activity of the serum, and then predicting the non-neutralized fraction of virus depending on the degree each epitope's activity is escaped and the contribution of each epitope to the total neutralizing activity[78]. These predictions were generated for the BF520 mutants used

in the neutralization assays depicted in Figure 3.4D and Figure 3.10. We chose to compare the fold change in IC80s because these values are similar to the level of neutralization we used in our deep mutational scanning selections.

We constrained the model for each antibody to have one epitope, while sera could have up to two epitopes. For figures in this manuscript and the interactive figures, we filter the mutations in the default view by requiring mutations to be present in at least three unique variants and to have a functional effect above -1.5. We filter mutations with low functional scores because variants with these mutations typically have low counts in no-antibody selections, which can cause high amounts of noise in their probability of escape scores. See https://dms-vep.github.io/HIV_Envelope_BF520_DMS_CD4bs_sera/ for interactive plots, notebooks detailing the fitting of these models, and PDBs with b-factors containing the escape values for each model.

3.8 Acknowledgements

We thank Michael Emerman for scientific advice. We thank Dr. Michel Nussenzweig and Marina Caskey for a gift of 3BNC117 IgGs. This work was supported in part by the NIH/NIAID under grants R01AI141707, R01AI140891, and U01AI169385 to JDB. This work was supported in part by a Gates Foundation grant INV-004949 to JDB. CER was supported by the Viral Pathogenesis and Evolution training grant T32 AI083203. KHDC was supported by NIH/NIAID grant F30AI149928. BD was funded in part by EMBO under grant ALTF 81-2020. TCY was supported by the CMB training grant T32 GM007270 and the NSF graduate research fellowship DGE-2140004. JDB is an Investigator of the Howard Hughes Medical Institute. PS is supported by the Emmy Noether Program of the German Research Foundation (DFG, Project No: 495793173). This research was supported by the Genomics & Bioinformatics Shared Resource (RRID:SCR_022606), Flow Cytometry Shared Resource (RRID:SCR_022613), and Scientific

Computing (NIH grants S10-OD-020069 and S10-OD-028685) of the Fred Hutch/University of Washington/Seattle Children's Cancer Consortium (P30 CA015704).

3.8.1 Competing Interests

JDB is on the scientific advisory boards of Apriori Bio, Invivyd, Aerium Therapeutics, and the Vaccine Company. JDB, KHDC, BD, ASD, and CER receive royalty payments as inventors on Fred Hutch licensed patents related to viral deep mutational scanning. ASD is currently an employee of Apriori Bio, though his contributions to this manuscript were performed when he was an employee of the Fred Hutch before he started work at Apriori Bio. A patent application encompassing aspects of this work has been filed by the University of Cologne and lists P.S. and F.K. as inventors. P.S., and F.K. received payments from the University of Cologne for licensed antibodies.

Chapter IV: Conclusions and future directions

4.1 Future directions for HIV Envelope deep mutational scanning

4.1.1 Expansion of HIV deep mutational scanning to more strains of Env

Expanding the use of the lentivirus deep mutational scanning system described here to more strains of HIV Envelope is an easily attainable step for future research. Since different strains of Env differ by dozens or more amino acids[9,30], the same mutations can have different effects depending which strain in which they are made[175,176,48,50]. Therefore, testing multiple strains of Env allows for a more complete picture of how a mutation may impact function or neutralization across HIV strains. The lentivirus deep mutational scanning system is amenable to pseudotyping with different HIV Env strains (Figure 3.2). Some strains produce higher titers of virus than others (Figure 3.2); this observation is the reason we chose to use the strain BF520 first rather than other potentially relevant strains such as BG505, which does not pseudotype the lentiviral vector well (Figure 3.2). Like with SARS-CoV-2 spike[150], modifications such as cytoplasmic tail deletions could potentially increase lentiviral pseudotype titers with some strains of Env, but this may require additional selections for incompletely understood phenotypes in the virus producing cells and can potentially alter Env antigenicity[177,178]. Therefore, unaltered Envs from strains commonly used in HIV research that pseudotype the lentiviral vector to high titers should be considered first. Two candidates are TRO.11 and CE1176, which pseudotype well (Figure 3.2) and are included in many studies since they are on a panel of viruses for evaluating antibody neutralization breadth against global Env diversity[166]. These Envs are from clade B and clade C, respectively, which will complement mutation effect data from clade A BF520 Env.

4.1.2 Evaluation of HIV vaccines and therapeutics

The lentiviral vector deep mutational scanning system described here is ready to be used to map escape from HIV therapeutics targeting Env. We have shown the system is capable of mapping escape from neutralization by antibodies and polyclonal serum. Many broadly neutralizing antibodies against Env are being tested as potential therapeutics and prophylactics alone and in combination[37,148]. We can map escape from neutralization by these antibodies in order to study potential synergistic or antagonistic effects of combinations. Likewise, escape can also be mapped for other protein or small molecule-based HIV entry inhibitors being developed as therapeutics against multi-drug resistant HIV[179,180]. We can use escape maps for these therapeutics across strains of HIV in combinations with knowledge of the geographic prevalence of each clade of HIV[30] to inform recommendations for therapeutic combinations for different populations.

In the future, this system can be used to evaluate experimental HIV vaccines. Experimental HIV vaccines that attempt to elicit broadly neutralizing antibodies have made significant progress over the years, but eliciting broad and potent neutralizing responses against fully glycosylated Envs remains difficult[39–41]. As a result, we were unable to find vaccine elicited serum capable of neutralizing BF520 Env to use with our mutant libraries. In Chapter 3, I described mapping of exceptionally broad and potent polyclonal sera elicited from natural infection and demonstrated how the lentivirus system is able to map neutralizing activity at multiple epitopes. The ability to finely map neutralization by serum at multiple epitopes will be valuable for informing vaccines by studying natural elicited sera, but also for evaluating future experimental vaccines, since immunity robust to HIV diversity and evolution will likely require neutralization at multiple epitopes[11,181]. By expanding the use of the lentivirus system to more HIV Env strains, including those on neutralization breadth panels and being used as immunogens, we can increase the likelihood of vaccine elicited sera being capable of

neutralizing one of our mutant libraries for mapping. The lentivirus system can then be used in combination with electron microscopy polyclonal epitope mapping to gain a more complete picture of serum neutralization and binding activity across different epitopes.

4.2 Potential uses of lentivirus deep mutational scanning beyond viral infection assays

While this dissertation described the use of the lentivirus deep mutational scanning system to measure the effects of mutations on virus entry into cells and escape from neutralizing antibodies and sera, the lentivirus deep mutational scanning system can be extended to measure the effects of mutations in more contexts. There are many steps in viral entry protein function that mutations can affect, including expression, receptor binding, and membrane fusion. Here, we modeled all of these as one phenotype; entry into cells. By measuring the effects of mutations at each step, we could study the relationship of each of these biophysical qualities with the additive latent and observed scales of the effects of mutation in models of epistasis[76,77]. There are many potential strategies for measuring the effects of mutation in different contexts, including cell surface selections for expression / conformational states with magnetic or fluorescence activated cell sorting[61,63], cell-to-cell fusion assays[61], incorporation of binding-deficient VSV-G to rescue fusion deficiencies and measure only receptor binding[182]. Like the virus infection assays used so far, each of these will require some degree of optimization. Similar assays are already being used to engineer immunogens more likely to bind desirable germline precursors for broadly neutralizing antibodies[41]. More complete characterization of the effects of mutations on protein functions alongside assays to identify mutations which increase binding to broadly neutralizing antibody precursors could aid in similar immunogen design.

4.3 Extension of lentivirus deep mutational scanning to other viruses

One of the major strengths of the lentivirus deep mutational scanning system is its ability to be used with any viral entry proteins capable of pseudotyping a lentivirus. This will allow it to be used to safely prospectively measure the effects of mutations on viral entry proteins for potential pandemic pathogens and evaluate therapeutics and vaccines being developed for them. Unfortunately, the system was not far along enough in its development to be used with SARS-CoV-2 spike immediately after the SARS-CoV-2 pandemic began, but some of the foundations of this system were used to rapidly develop a SARS-CoV-2 pseudovirus neutralization assay using lentiviruses[70]. In the future, because of the work described here, adapting the system to use with new viral entry proteins will take months rather than years. Building a knowledge base of the effects of mutations on different functions of potential pandemic viral entry proteins can aid in monitoring of viruses, evaluate therapeutics for potential virus escape, and could be critical for rapidly designing vaccines, as with SARS-CoV-2. Despite the clear benefits of these studies and the safety of the lentivirus system, data from this system could potentially be used to engineer gain of function mutations into potential pandemic pathogens in separate downstream experiments, highlighting the need for frameworks to rigorously evaluate the public health cost-benefit tradeoff of the accessibility of these data and their use in any downstream experiments[122].

References

1. White, J.M., Delos, S.E., Brecher, M., and Schornberg, K. (2008). Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme. Crit. Rev. Biochem. Mol. Biol. *43*, 189–219. 10.1080/10409230802058320.

2. Murin, C.D., Wilson, I.A., and Ward, A.B. (2019). Antibody responses to viral infections: a structural perspective across three different enveloped viruses. Nat. Microbiol. *4*, 734–747. 10.1038/s41564-019-0392-y.

3. Wu, N.C., and Wilson, I.A. (2020). Influenza Hemagglutinin Structures and Antibody Recognition. Cold Spring Harb. Perspect. Med. *10*, a038778. 10.1101/cshperspect.a038778.

4. Burton, D.R., and Hangartner, L. (2016). Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. Annu. Rev. Immunol. *34*, 635–659. 10.1146/annurev-immunol-041015-055515.

5. Crispin, M., Ward, A.B., and Wilson, I.A. (2018). Structure and Immune Recognition of the HIV Glycan Shield. Annu. Rev. Biophys. *47*, 499–523. 10.1146/annurev-biophys-060414-034156.

6. Altman, M.O., Angel, M., Košík, I., Trovão, N.S., Zost, S.J., Gibbs, J.S., Casalino, L., Amaro, R.E., Hensley, S.E., Nelson, M.I., et al. (2019). Human Influenza A Virus Hemagglutinin Glycan Evolution Follows a Temporal Pattern to a Glycan Limit. mBio *10*, e00204-19. 10.1128/mBio.00204-19.

7. Volz, E.M., Koelle, K., and Bedford, T. (2013). Viral Phylodynamics. PLOS Comput. Biol. *9*, e1002947. 10.1371/journal.pcbi.1002947.

8. Kistler, K.E., Huddleston, J., and Bedford, T. (2022). Rapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2. Cell Host Microbe *30*, 545-555.e4. 10.1016/j.chom.2022.03.018.

9. Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., and Detours, V. (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. Br. Med. Bull. *58*, 19–42. 10.1093/bmb/58.1.19.

10. Muñoz-Alía, M.Á., Nace, R.A., Zhang, L., and Russell, S.J. (2021). Serotypic evolution of measles virus is constrained by multiple co-dominant B cell epitopes on its surface glycoproteins. Cell Rep. Med. *2*, 100225. 10.1016/j.xcrm.2021.100225.

11. Greaney, A.J., Welsh, F.C., and Bloom, J.D. (2021). Co-dominant neutralizing epitopes make anti-measles immunity resistant to viral evolution. Cell Rep. Med. *2*, 100257. 10.1016/j.xcrm.2021.100257.

12. Lee, J.M., Eguia, R., Zost, S.J., Choudhary, S., Wilson, P.C., Bedford, T., Stevens-Ayers, T., Boeckh, M., Hurt, A.C., Lakdawala, S.S., et al. (2019). Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. eLife *8*, e49324. 10.7554/eLife.49324.

13. Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.-L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science *367*, 1260–1263. 10.1126/science.abb2507.

14. Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell *181*, 281-292.e6. 10.1016/j.cell.2020.02.058.

15. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell *181*, 271-280.e8. 10.1016/j.cell.2020.02.052.

16. Pallesen, J., Wang, N., Corbett, K.S., Wrapp, D., Kirchdoerfer, R.N., Turner, H.L., Cottrell, C.A., Becker, M.M., Wang, L., Shi, W., et al. (2017). Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. Proc. Natl. Acad. Sci. *114*, E7348–E7357. 10.1073/pnas.1707304114.

17. Dai, L., and Gao, G.F. (2021). Viral targets for vaccines against COVID-19. Nat. Rev. Immunol. *21*, 73–82. 10.1038/s41577-020-00480-0.

18. Golob, J.L., Lugogo, N., Lauring, A.S., and Lok, A.S. SARS-CoV-2 vaccines: a triumph of science and collaboration. JCI Insight *6*, e149187. 10.1172/jci.insight.149187.

19. Li, Y., Myers, J.L., Bostick, D.L., Sullivan, C.B., Madara, J., Linderman, S.L., Liu, Q., Carter, D.M., Wrammert, J., Esposito, S., et al. (2013). Immune history shapes specificity of pandemic H1N1 influenza antibody responses. J. Exp. Med. *210*, 1493–1500. 10.1084/jem.20130212.

20. Huang, K.-Y.A., Rijal, P., Schimanski, L., Powell, T.J., Lin, T.-Y., McCauley, J.W., Daniels, R.S., and Townsend, A.R. (2015). Focused antibody response to influenza linked to antigenic drift. J. Clin. Invest. *125*, 2631–2645. 10.1172/JCI81104.

21. Davis, A.K.F., McCormick, K., Gumina, M.E., Petrie, J.G., Martin, E.T., Xue, K.S., Bloom, J.D., Monto, A.S., Bushman, F.D., and Hensley, S.E. (2018). Sera from Individuals with Narrowly Focused Influenza Virus Antibodies Rapidly Select Viral Escape Mutations In Ovo. J. Virol. *92*, e00859-18. 10.1128/JVI.00859-18.

22. Weisblum, Y., Schmidt, F., Zhang, F., DaSilva, J., Poston, D., Lorenzi, J.C., Muecksch, F., Rutkowska, M., Hoffmann, H.-H., Michailidis, E., et al. (2020). Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. eLife *9*, e61312. 10.7554/eLife.61312.

23. Greaney, A.J., Loes, A.N., Crawford, K.H.D., Starr, T.N., Malone, K.D., Chu, H.Y., and Bloom, J.D. (2021). 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. 10.1016/j.chom.2021.02.003.

24. Greaney, A.J., Starr, T.N., Barnes, C.O., Weisblum, Y., Schmidt, F., Caskey, M., Gaebler, C., Cho, A., Agudelo, M., Finkin, S., et al. (2021). Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. Nat. Commun. *12*, 4196. 10.1038/s41467-021-24435-8.

25. Chan, D.C., Fass, D., Berger, J.M., and Kim, P.S. (1997). Core Structure of gp41 from the HIV Envelope Glycoprotein. Cell *89*, 263–273. 10.1016/S0092-8674(00)80205-6.

26. Sattentau, Q.J., and Weiss, R.A. (1988). The CD4 antigen: Physiological ligand and HIV receptor. Cell *52*, 631–633. 10.1016/0092-8674(88)90397-2.

27. Lu, M., Ma, X., Castillo-Menendez, L.R., Gorman, J., Alsahafi, N., Ermel, U., Terry, D.S., Chambers, M., Peng, D., Zhang, B., et al. (2019). Associating HIV-1 envelope glycoprotein structures with states on the virus observed by smFRET. Nature *568*, 415–419. 10.1038/s41586-019-1101-y.

28. Chen, B. (2019). Molecular Mechanism of HIV-1 Entry. Trends Microbiol. *27*, 878–891. 10.1016/j.tim.2019.06.002.

29. Kwong, P.D., Doyle, M.L., Casper, D.J., Cicala, C., Leavitt, S.A., Majeed, S., Steenbeke, T.D., Venturi, M., Chaiken, I., Fung, M., et al. (2002). HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. Nature *420*, 678–682. 10.1038/nature01188.

30. Fischer, W., Giorgi, E.E., Chakraborty, S., Nguyen, K., Bhattacharya, T., Theiler, J., Goloboff, P.A., Yoon, H., Abfalterer, W., Foley, B.T., et al. (2021). HIV-1 and SARS-CoV-2: Patterns in the evolution of two pandemic pathogens. Cell Host Microbe *29*, 1093–1110. 10.1016/j.chom.2021.05.012.

31. Preston, B.D., Poiesz, B.J., and Loeb, L.A. (1988). Fidelity of HIV-1 Reverse Transcriptase. Science *242*, 1168–1171. 10.1126/science.2460924.

32. Rambaut, A., Posada, D., Crandall, K.A., and Holmes, E.C. (2004). The causes and consequences of HIV evolution. Nat. Rev. Genet. *5*, 52–61. 10.1038/nrg1246.

33. Richman, D.D., Wrin, T., Little, S.J., and Petropoulos, C.J. (2003). Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc. Natl. Acad. Sci. *100*, 4144–4149. 10.1073/pnas.0630530100.

34. Pegu, A., Hessell, A.J., Mascola, J.R., and Haigwood, N.L. (2017). Use of broadly neutralizing antibodies for HIV-1 prevention. Immunol. Rev. *275*, 296–312. 10.1111/imr.12511.

35. Stephenson, K.E., Wagh, K., Korber, B., and Barouch, D.H. (2020). Vaccines and Broadly Neutralizing Antibodies for HIV-1 Prevention. Annu. Rev. Immunol. *38*, 673–703. 10.1146/annurev-immunol-080219-023629.

36. Sok, D., and Burton, D.R. (2018). Recent progress in broadly neutralizing antibodies to HIV. Nat. Immunol. *19*, 1179–1188. 10.1038/s41590-018-0235-7.

37. Gruell, H., and Schommers, P. (2022). Broadly neutralizing antibodies against HIV-1 and concepts for application. Curr. Opin. Virol. *54*, 101211. 10.1016/j.coviro.2022.101211.

38. Haynes, B.F., Wiehe, K., Borrrow, P., Saunders, K.O., Korber, B., Wagh, K., McMichael, A.J., Kelsoe, G., Hahn, B.H., Alt, F., et al. (2023). Strategies for HIV-1 vaccines that induce

broadly neutralizing antibodies. Nat. Rev. Immunol. *23*, 142–158. 10.1038/s41577-022-00753-w.

39. Xu, K., Acharya, P., Kong, R., Cheng, C., Chuang, G.-Y., Liu, K., Louder, M.K., O'Dell, S., Rawi, R., Sastry, M., et al. (2018). Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1. Nat. Med. *24*, 857–867. 10.1038/s41591-018-0042-6.

40. Dubrovskaya, V., Tran, K., Ozorowski, G., Guenaga, J., Wilson, R., Bale, S., Cottrell, C.A., Turner, H.L., Seabright, G., O'Dell, S., et al. (2019). Vaccination with Glycan-Modified HIV NFL Envelope Trimer-Liposomes Elicits Broadly Neutralizing Antibodies to Multiple Sites of Vulnerability. Immunity *51*, 915-929.e7. 10.1016/j.immuni.2019.10.008.

41. Gristick, H.B., Hartweger, H., Loewe, M., van Schooten, J., Ramos, V., Oliveira, T.Y., Nishimura, Y., Koranda, N.S., Wall, A., Yao, K.-H., et al. (2023). CD4 binding site immunogens elicit heterologous anti–HIV-1 neutralizing antibodies in transgenic and wild-type animals. Sci. Immunol. *8*, eade6364. 10.1126/sciimmunol.ade6364.

42. Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807. 10.1038/nmeth.3027.

43. Tenthorey, J.L., Emerman, M., and Malik, H.S. (2022). Evolutionary Landscapes of Host-Virus Arms Races. Annu. Rev. Immunol. *40*, 271–294. 10.1146/annurev-immunol-072621-084422.

44. Narayanan, K.K., and Procko, E. (2021). Deep Mutational Scanning of Viral Glycoproteins and Their Host Receptors. Front. Mol. Biosci. *8*.

45. Bloom, J.D. (2014). An Experimentally Determined Evolutionary Model Dramatically Improves Phylogenetic Fit. Mol. Biol. Evol. *31*, 1956–1978. 10.1093/molbev/msu173.

46. Doud, M.B., Hensley, S.E., and Bloom, J.D. (2017). Complete mapping of viral escape from neutralizing antibodies. PLOS Pathog. *13*, e1006271. 10.1371/journal.ppat.1006271.

47. Haddox, H.K., Dingens, A.S., and Bloom, J.D. (2016). Experimental Estimation of the Effects of All Amino-Acid Mutations to HIV's Envelope Protein on Viral Replication in Cell Culture. PLOS Pathog. *12*, e1006114. 10.1371/journal.ppat.1006114.

48. Haddox, H.K., Dingens, A.S., Hilton, S.K., Overbaugh, J., and Bloom, J.D. (2018). Mapping mutational effects along the evolutionary landscape of HIV envelope. eLife *7*, e34420. 10.7554/eLife.34420.

49. Dingens, A.S., Haddox, H.K., Overbaugh, J., and Bloom, J.D. (2017). Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody. Cell Host Microbe *21*, 777-787.e4. 10.1016/j.chom.2017.05.003.

50. Dingens, A.S., Arenz, D., Weight, H., Overbaugh, J., and Bloom, J.D. (2019). An Antigenic Atlas of HIV-1 Escape from Broadly Neutralizing Antibodies Distinguishes Functional and Structural Epitopes. Immunity *50*, 520-532.e3. 10.1016/j.immuni.2018.12.017.

51. Dingens, A.S., Pratap, P., Malone, K., Hilton, S.K., Ketas, T., Cottrell, C.A., Overbaugh, J., Moore, J.P., Klasse, P., Ward, A.B., et al. (2021). High-resolution mapping of the neutralizing and binding specificities of polyclonal sera post-HIV Env trimer vaccination. eLife *10*, e64281. 10.7554/eLife.64281.

52. Garrett, M.E., Itell, H.L., Crawford, K.H.D., Basom, R., Bloom, J.D., and Overbaugh, J. (2020). Phage-DMS: A Comprehensive Method for Fine Mapping of Antibody Epitopes. iScience *23*, 101622. 10.1016/j.isci.2020.101622.

53. Simonich, C., Shipley, M.M., Doepker, L., Gobillot, T., Garrett, M., Cale, E.M., Hennessy, B., Itell, H., Chohan, V., Doria-Rose, N., et al. (2021). A diverse collection of B cells responded to HIV infection in infant BG505. Cell Rep. Med. *2*, 100314. 10.1016/j.xcrm.2021.100314.

54. Stoddard, C.I., Galloway, J., Chu, H.Y., Shipley, M.M., Sung, K., Itell, H.L., Wolf, C.R., Logue, J.K., Magedson, A., Garrett, M.E., et al. (2021). Epitope profiling reveals binding signatures of SARS-CoV-2 immune response in natural infection and cross-reactivity with endemic human CoVs. Cell Rep. *35*, 109164. 10.1016/j.celrep.2021.109164.

55. Tiu, C.K., Zhu, F., Wang, L.-F., and de Alwis, R. (2022). Phage ImmunoPrecipitation Sequencing (PhIP-Seq): The Promise of High Throughput Serology. Pathogens *11*, 568. 10.3390/pathogens11050568.

56. Garrett, M.E., Galloway, J.G., Wolf, C., Logue, J.K., Franko, N., Chu, H.Y., Matsen, F.A., IV, and Overbaugh, J.M. (2022). Comprehensive characterization of the antibody responses to SARS-CoV-2 Spike protein finds additional vaccine-induced epitopes beyond those for mild infection. eLife *11*, e73490. 10.7554/eLife.73490.

57. Starr, T.N., Greaney, A.J., Hilton, S.K., Ellis, D., Crawford, K.H.D., Dingens, A.S., Navarro, M.J., Bowen, J.E., Tortorici, M.A., Walls, A.C., et al. (2020). Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell *182*, 1295-1310.e20. 10.1016/j.cell.2020.08.012.

58. Starr, T.N., Greaney, A.J., Addetia, A., Hannon, W.W., Choudhary, M.C., Dingens, A.S., Li, J.Z., and Bloom, J.D. (2021). Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. Science *371*, 850–854. 10.1126/science.abf9302.

59. Greaney, A.J., Starr, T.N., Gilchuk, P., Zost, S.J., Binshtein, E., Loes, A.N., Hilton, S.K., Huddleston, J., Eguia, R., Crawford, K.H.D., et al. (2021). Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. Cell Host Microbe *29*, 44-57.e9. 10.1016/j.chom.2020.11.007.

60. Starr, T.N., Zepeda, S.K., Walls, A.C., Greaney, A.J., Alkhovsky, S., Veesler, D., and Bloom, J.D. (2022). ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. Nature *603*, 913–918. 10.1038/s41586-022-04464-z.

61. Salamango, D.J., Alam, K.K., Burke, D.H., and Johnson, M.C. (2016). In Vivo Analysis of Infectivity, Fusogenicity, and Incorporation of a Mutagenic Viral Glycoprotein Library Reveals Determinants for Virus Incorporation. J. Virol. *90*, 6502–6514. 10.1128/JVI.00804-16.

62. Steichen, J.M., Kulp, D.W., Tokatlian, T., Escolano, A., Dosenovic, P., Stanfield, R.L., McCoy, L.E., Ozorowski, G., Hu, X., Kalyuzhniy, O., et al. (2016). HIV Vaccine Design to Target Germline Precursors of Glycan-Dependent Broadly Neutralizing Antibodies. Immunity *45*, 483–496. 10.1016/j.immuni.2016.08.016.

63. Tan, T.J.C., Mou, Z., Lei, R., Ouyang, W.O., Yuan, M., Song, G., Andrabi, R., Wilson, I.A., Kieffer, C., Dai, X., et al. (2023). High-throughput identification of prefusion-stabilizing mutations in SARS-CoV-2 spike. Nat. Commun. *14*, 2003. 10.1038/s41467-023-37786-1.

64. Bianchi, M., Turner, H.L., Nogal, B., Cottrell, C.A., Oyen, D., Pauthner, M., Bastidas, R., Nedellec, R., McCoy, L.E., Wilson, I.A., et al. (2018). Electron-Microscopy-Based Epitope Mapping Defines Specificities of Polyclonal Antibodies Elicited during HIV-1 BG505 Envelope Trimer Immunization. Immunity *49*, 288-300.e8. 10.1016/j.immuni.2018.07.009.

65. Nogal, B., Bianchi, M., Cottrell, C.A., Kirchdoerfer, R.N., Sewall, L.M., Turner, H.L., Zhao, F., Sok, D., Burton, D.R., Hangartner, L., et al. (2020). Mapping Polyclonal Antibody Responses in Non-human Primates Vaccinated with HIV Env Trimer Subunit Vaccines. Cell Rep. *30*, 3755-3765.e7. 10.1016/j.celrep.2020.02.061.

66. Antanasijevic, A., Sewall, L.M., Cottrell, C.A., Carnathan, D.G., Jimenez, L.E., Ngo, J.T., Silverman, J.B., Groschel, B., Georgeson, E., Bhiman, J., et al. (2021). Polyclonal antibody responses to HIV Env immunogens resolved using cryoEM. Nat. Commun. *12*, 4817. 10.1038/s41467-021-25087-4.

67. Zufferey, R., Dull, T., Mandel, R.J., Bukovsky, A., Quiroz, D., Naldini, L., and Trono, D. (1998). Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. J. Virol. *72*, 9873–9880. 10.1128/JVI.72.12.9873-9880.1998.

68. Cronin, J., Zhang, X.-Y., and Reiser, J. (2005). Altering the tropism of lentiviral vectors through pseudotyping. Curr. Gene Ther. *5*, 387–398. 10.2174/1566523054546224.

69. Jorgenson, R.L., Vogt, V.M., and Johnson, M.C. (2009). Foreign Glycoproteins Can Be Actively Recruited to Virus Assembly Sites during Pseudotyping. J. Virol. *83*, 4060–4067. 10.1128/JVI.02425-08.

70. Crawford, K.H.D., Eguia, R., Dingens, A.S., Loes, A.N., Malone, K.D., Wolf, C.R., Chu, H.Y., Tortorici, M.A., Veesler, D., Murphy, M., et al. (2020). Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays. Viruses *12*, 513. 10.3390/v12050513.

71. Hill, A.J., McFaline-Figueroa, J.L., Starita, L.M., Gasperini, M.J., Matreyek, K.A., Packer, J., Jackson, D., Shendure, J., and Trapnell, C. (2018). On the design of CRISPR-based single cell molecular screens. Nat. Methods *15*, 271–274. 10.1038/nmeth.4604.

72. Schlub, T.E., Smyth, R.P., Grimm, A.J., Mak, J., and Davenport, M.P. (2010). Accurately Measuring Recombination between Closely Related HIV-1 Genomes. PLOS Comput. Biol. *6*, e1000766. 10.1371/journal.pcbi.1000766.

73. Jetzt, A.E., Yu, H., Klarmann, G.J., Ron, Y., Preston, B.D., and Dougherty, J.P. (2000). High Rate of Recombination throughout the Human Immunodeficiency Virus Type 1 Genome. J. Virol. *74*, 1234–1240.

74. Starr, T.N., and Thornton, J.W. (2016). Epistasis in protein evolution. Protein Sci. *25*, 1204–1218. 10.1002/pro.2897.

75. Sailer, Z.R., and Harms, M.J. (2017). Molecular ensembles make evolution unpredictable. Proc. Natl. Acad. Sci. *114*, 11938–11943. 10.1073/pnas.1711927114.

76. Sailer, Z.R., and Harms, M.J. (2017). Detecting High-Order Epistasis in Nonlinear Genotype-Phenotype Maps. Genetics *205*, 1079–1088. 10.1534/genetics.116.195214.

77. Otwinowski, J., McCandlish, D.M., and Plotkin, J.B. (2018). Inferring the shape of global epistasis. Proc. Natl. Acad. Sci. *115*, E7550–E7558. 10.1073/pnas.1804015115.

78. Yu, T.C., Thornton, Z.T., Hannon, W.W., DeWitt, W.S., Radford, C.E., Matsen, F.A., IV, and Bloom, J.D. (2022). A biophysical model of viral escape from polyclonal antibodies. Virus Evol. *8*, veac110. 10.1093/ve/veac110.

79. Cao, Y., Jian, F., Wang, J., Yu, Y., Song, W., Yisimayi, A., Wang, J., An, R., Chen, X., Zhang, N., et al. (2023). Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. Nature *614*, 521–529. 10.1038/s41586-022-05644-7.

80. Liu, L., Iketani, S., Guo, Y., Chan, J.F.-W., Wang, M., Liu, L., Luo, Y., Chu, H., Huang, Y., Nair, M.S., et al. (2022). Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. Nature *602*, 676–681. 10.1038/s41586-021-04388-0.

81. Wang, Q., Guo, Y., Iketani, S., Nair, M.S., Li, Z., Mohri, H., Wang, M., Yu, J., Bowen, A.D., Chang, J.Y., et al. (2022). Antibody evasion by SARS-CoV-2 Omicron subvariants BA.2.12.1, BA.4 and BA.5. Nature *608*, 603–608. 10.1038/s41586-022-05053-w.

82. Wang, Q., Iketani, S., Li, Z., Guo, Y., Yeh, A.Y., Liu, M., Yu, J., Sheng, Z., Huang, Y., Liu, L., et al. (2022). Antigenic characterization of the SARS-CoV-2 Omicron subvariant BA.2.75. Cell Host Microbe *30*, 1512-1517.e4. 10.1016/j.chom.2022.09.002.

83. Cao, Y., Jian, F., Wang, J., Yu, Y., Song, W., Yisimayi, A., Wang, J., An, R., Chen, X., Zhang, N., et al. (2022). Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. Nature. 10.1038/s41586-022-05644-7.

84. Starr, T.N., Czudnochowski, N., Liu, Z., Zatta, F., Park, Y.-J., Addetia, A., Pinto, D., Beltramello, M., Hernandez, P., Greaney, A.J., et al. (2021). SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. Nature *597*, 97–102. 10.1038/s41586-021-03807-6.

85. 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 Evol. *8*, veac021. 10.1093/ve/veac021.

86. Ouyang, W.O., Tan, T.J.C., Lei, R., Song, G., Kieffer, C., Andrabi, R., Matreyek, K.A., and Wu, N.C. (2022). Probing the biophysical constraints of SARS-CoV-2 spike N-terminal domain using deep mutational scanning. Sci. Adv. *8*, eadd7221. 10.1126/sciadv.add7221.

87. Javanmardi, K., Chou, C.-W., Terrace, C.I., Annapareddy, A., Kaoud, T.S., Guo, Q., Lutgens, J., Zorkic, H., Horton, A.P., Gardner, E.C., et al. (2021). Rapid characterization of

spike variants via mammalian cell surface display. Mol. Cell *81*, 5099-5111.e8. 10.1016/j.molcel.2021.11.024.

88. Starr, T.N., Greaney, A.J., Stewart, C.M., Walls, A.C., Hannon, W.W., Veesler, D., and Bloom, J.D. (2022). Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA.1 and BA.2 receptor-binding domains. PLOS Pathog. *18*, e1010951. 10.1371/journal.ppat.1010951.

89. Gilbert, P.B., Montefiori, D.C., McDermott, A.B., Fong, Y., Benkeser, D., Deng, W., Zhou, H., Houchens, C.R., Martins, K., Jayashankar, L., et al. (2022). Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. Science *375*, 43–50. 10.1126/science.abm3425.

90. Feng, S., Phillips, D.J., White, T., Sayal, H., Aley, P.K., Bibi, S., Dold, C., Fuskova, M., Gilbert, S.C., Hirsch, I., et al. (2021). Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. Nat. Med. *27*, 2032–2040. 10.1038/s41591-021-01540-1.

91. Naldini, L., Blömer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M., and Trono, D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science *272*, 263–267. 10.1126/science.272.5259.263.

92. OhAinle, M., Helms, L., Vermeire, J., Roesch, F., Humes, D., Basom, R., Delrow, J.J., Overbaugh, J., and Emerman, M. (2018). A virus-packageable CRISPR screen identifies host factors mediating interferon inhibition of HIV. eLife *7*, e39823. 10.7554/eLife.39823.

93. Zhao, X., Zheng, S., Chen, D., Zheng, M., Li, X., Li, G., Lin, H., Chang, J., Zeng, H., and Guo, J.-T. (2020). LY6E Restricts Entry of Human Coronaviruses, Including Currently Pandemic SARS-CoV-2. J. Virol. *94*, e00562-20. 10.1128/JVI.00562-20.

94. Peacock, T.P., Goldhill, D.H., Zhou, J., Baillon, L., Frise, R., Swann, O.C., Kugathasan, R., Penn, R., Brown, J.C., Sanchez-David, R.Y., et al. (2021). The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. Nat. Microbiol. *6*, 899–909. 10.1038/s41564-021-00908-w.

95. Zheng, M., Zhao, X., Zheng, S., Chen, D., Du, P., Li, X., Jiang, D., Guo, J.-T., Zeng, H., and Lin, H. (2020). Bat SARS-Like WIV1 coronavirus uses the ACE2 of multiple animal species as receptor and evades IFITM3 restriction via TMPRSS2 activation of membrane fusion. Emerg. Microbes Infect. *9*, 1567–1579. 10.1080/22221751.2020.1787797.

96. Khare, S., Gurry, C., Freitas, L., Schultz, M.B., Bach, G., Diallo, A., Akite, N., Ho, J., Lee, R.T., Yeo, W., et al. (2021). GISAID's Role in Pandemic Response. China CDC Wkly. *3*, 1049–1051. 10.46234/ccdcw2021.255.

97. Turakhia, Y., Thornlow, B., Hinrichs, A.S., De Maio, N., Gozashti, L., Lanfear, R., Haussler, D., and Corbett-Detig, R. (2021). Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. Nat. Genet. *53*, 809–816. 10.1038/s41588-021-00862-7.

98. Maher, M.C., Bartha, I., Weaver, S., di Iulio, J., Ferri, E., Soriaga, L., Lempp, F.A., Hie, B.L., Bryson, B., Berger, B., et al. (2022). Predicting the mutational drivers of future SARS-

CoV-2 variants of concern. Sci. Transl. Med. *14*, eabk3445. 10.1126/scitranslmed.abk3445.

99. Hiatt, J.B., Patwardhan, R.P., Turner, E.H., Lee, C., and Shendure, J. (2010). Parallel, tag-directed assembly of locally derived short sequence reads. Nat. Methods *7*, 119–122. 10.1038/nmeth.1416.

100. Matreyek, K.A., Starita, L.M., Stephany, J.J., Martin, B., Chiasson, M.A., Gray, V.E., Kircher, M., Khechaduri, A., Dines, J.N., Hause, R.J., et al. (2018). Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat. Genet. *50*, 874–882. 10.1038/s41588-018-0122-z.

101. Cerutti, G., Guo, Y., Wang, P., Nair, M.S., Wang, M., Huang, Y., Yu, J., Liu, L., Katsamba, P.S., Bahna, F., et al. (2021). Neutralizing antibody 5-7 defines a distinct site of vulnerability in SARS-CoV-2 spike N-terminal domain. Cell Rep. *37*. 10.1016/j.celrep.2021.109928.

102. Westendorf, K., Žentelis, S., Wang, L., Foster, D., Vaillancourt, P., Wiggin, M., Lovett, E., Lee, R. van der, Hendle, J., Pustilnik, A., et al. (2022). LY-CoV1404 (bebtelovimab) potently neutralizes SARS-CoV-2 variants. Cell Rep. *39*. 10.1016/j.celrep.2022.110812.

103. Zhou, P., Song, G., Liu, H., Yuan, M., He, W., Beutler, N., Zhu, X., Tse, L.V., Martinez, D.R., Schäfer, A., et al. (2023). Broadly neutralizing anti-S2 antibodies protect against all three human betacoronaviruses that cause deadly disease. Immunity *56*, 669-686.e7. 10.1016/j.immuni.2023.02.005.

104. McCallum, M., De Marco, A., Lempp, F.A., Tortorici, M.A., Pinto, D., Walls, A.C., Beltramello, M., Chen, A., Liu, Z., Zatta, F., et al. (2021). N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell *184*, 2332-2347.e16. 10.1016/j.cell.2021.03.028.

105. Chen, C., Nadeau, S., Yared, M., Voinov, P., Xie, N., Roemer, C., and Stadler, T. (2022). CoV-Spectrum: analysis of globally shared SARS-CoV-2 data to identify and characterize new variants. Bioinformatics *38*, 1735–1737. 10.1093/bioinformatics/btab856.

106. Starr, T.N., Greaney, A.J., Stewart, C.M., Walls, A.C., Hannon, W.W., Veesler, D., and Bloom, J.D. (2022). Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA.1 and BA.2 receptor-binding domains. PLOS Pathog. *18*, e1010951. 10.1371/journal.ppat.1010951.

107. Baum, A., Fulton, B.O., Wloga, E., Copin, R., Pascal, K.E., Russo, V., Giordano, S., Lanza, K., Negron, N., Ni, M., et al. (2020). Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science *369*, 1014–1018. 10.1126/science.abd0831.

108. Hansen, J., Baum, A., Pascal, K.E., Russo, V., Giordano, S., Wloga, E., Fulton, B.O., Yan, Y., Koon, K., Patel, K., et al. (2020). Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. Science *369*, 1010–1014. 10.1126/science.abd0827.

109. Bloom, J.D., Beichman, A.C., Neher, R.A., and Harris, K. (2023). Evolution of the SARS-CoV-2 Mutational Spectrum. Mol. Biol. Evol. *40*, msad085. 10.1093/molbev/msad085.

110. Zhang, J., Cai, Y., Xiao, T., Lu, J., Peng, H., Sterling, S.M., Walsh, R.M., Rits-Volloch, S., Zhu, H., Woosley, A.N., et al. (2021). Structural impact on SARS-CoV-2 spike protein by D614G substitution. Science *372*, 525–530. 10.1126/science.abf2303.

111. Benton, D.J., Wrobel, A.G., Roustan, C., Borg, A., Xu, P., Martin, S.R., Rosenthal, P.B., Skehel, J.J., and Gamblin, S.J. (2021). The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2. Proc. Natl. Acad. Sci. *118*, e2022586118. 10.1073/pnas.2022586118.

112. Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., et al. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. Nature *592*, 116–121. 10.1038/s41586-020-2895-3.

113. Bellusci, L., Grubbs, G., Zahra, F.T., Forgacs, D., Golding, H., Ross, T.M., and Khurana, S. (2022). Antibody affinity and cross-variant neutralization of SARS-CoV-2 Omicron BA.1, BA.2 and BA.3 following third mRNA vaccination. Nat. Commun. *13*, 4617. 10.1038/s41467-022-32298-w.

114. DeGrace, M.M., Ghedin, E., Frieman, M.B., Krammer, F., Grifoni, A., Alisoltani, A., Alter, G., Amara, R.R., Baric, R.S., Barouch, D.H., et al. (2022). Defining the risk of SARS-CoV-2 variants on immune protection. Nature *605*, 640–652. 10.1038/s41586-022-04690-5.

115. Ballal, A., Malliaris, C.D., and Morozov, A.V. (2020). Molecular Phenotypes as Key Intermediates in Mapping Genotypes to Fitness. In Evolutionary Biology—A Transdisciplinary Approach, P. Pontarotti, ed. (Springer International Publishing), pp. 15–40. 10.1007/978-3-030-57246-4_2.

116. Harms, M.J., and Thornton, J.W. (2013). Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nat. Rev. Genet. *14*, 559–571. 10.1038/nrg3540.

117. Huang, S.-W., Tai, C.-H., Hsu, Y.-M., Cheng, D., Hung, S.-J., Chai, K.M., Wang, Y.-F., and Wang, J.-R. (2020). Assessing the application of a pseudovirus system for emerging SARS-CoV-2 and re-emerging avian influenza virus H5 subtypes in vaccine development. Biomed. J. *43*, 375–387. 10.1016/j.bj.2020.06.003.

118. Khetawat, D., and Broder, C.C. (2010). A Functional Henipavirus Envelope Glycoprotein Pseudotyped Lentivirus Assay System. Virol. J. *7*, 312. 10.1186/1743-422X-7-312.

119. Kobinger, G.P., Weiner, D.J., Yu, Q.-C., and Wilson, J.M. (2001). Filovirus-pseudotyped lentiviral vector can efficiently and stably transduce airway epithelia in vivo. Nat. Biotechnol. *19*, 225–230. 10.1038/85664.

120. Medina, M.F., Kobinger, G.P., Rux, J., Gasmi, M., Looney, D.J., Bates, P., and Wilson, J.M. (2003). Lentiviral vectors pseudotyped with minimal filovirus envelopes increased gene transfer in murine lung. Mol. Ther. *8*, 777–789. 10.1016/j.ymthe.2003.07.003.

121. Larson, R.A., Dai, D., Hosack, V.T., Tan, Y., Bolken, T.C., Hruby, D.E., and Amberg, S.M. (2008). Identification of a Broad-Spectrum Arenavirus Entry Inhibitor. J. Virol. *82*, 10768–10775. 10.1128/JVI.00941-08.

122. Pannu, J., Palmer, M.J., Cicero, A., Relman, D.A., Lipsitch, M., and Inglesby, T. (2022). Strengthen oversight of risky research on pathogens. Science *378*, 1170–1172. 10.1126/science.adf6020.

123. Andreano, E., Piccini, G., Licastro, D., Casalino, L., Johnson, N.V., Paciello, I., Dal Monego, S., Pantano, E., Manganaro, N., Manenti, A., et al. (2021). SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma. Proc. Natl. Acad. Sci. *118*, e2103154118. 10.1073/pnas.2103154118.

124. Crawford, K.H.D., Dingens, A.S., Eguia, R., Wolf, C.R., Wilcox, N., Logue, J.K., Shuey, K., Casto, A.M., Fiala, B., Wrenn, S., et al. (2021). Dynamics of Neutralizing Antibody Titers in the Months After Severe Acute Respiratory Syndrome Coronavirus 2 Infection. J. Infect. Dis. *223*, 197–205. 10.1093/infdis/jiaa618.

125. Havranek, K.E., Jimenez, A.R., Acciani, M.D., Lay Mendoza, M.F., Reyes Ballista, J.M., Diaz, D.A., and Brindley, M.A. (2020). SARS-CoV-2 Spike Alterations Enhance Pseudoparticle Titers and Replication-Competent VSV-SARS-CoV-2 Virus. Viruses *12*, 1465. 10.3390/v12121465.

126. McCarthy, K.R., Rennick, L.J., Nambulli, S., Robinson-McCarthy, L.R., Bain, W.G., Haidar, G., and Duprex, W.P. (2021). Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science *371*, 1139–1142. 10.1126/science.abf6950.

127. Dingens, A.S., Acharya, P., Haddox, H.K., Rawi, R., Xu, K., Chuang, G.-Y., Wei, H., Zhang, B., Mascola, J.R., Carragher, B., et al. (2018). Complete functional mapping of infection- and vaccine-elicited antibodies against the fusion peptide of HIV. PLOS Pathog. *14*, e1007159. 10.1371/journal.ppat.1007159.

128. Chun, T.-W., Carruth, L., Finzi, D., Shen, X., DiGiuseppe, J.A., Taylor, H., Hermankova, M., Chadwick, K., Margolick, J., Quinn, T.C., et al. (1997). Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature *387*, 183–188. 10.1038/387183a0.

129. Pang, S., Koyanagi, Y., Miles, S., Wiley, C., Vinters, H.V., and Chen, I.S.Y. (1990). High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. Nature *343*, 85–89. 10.1038/343085a0.

130. Sharkey, M.E., Teo, I., Greenough, T., Sharova, N., Luzuriaga, K., Sullivan, J.L., Bucy, R.P., Kostrikis, L.G., Haase, A., Veryard, C., et al. (2000). Persistence of episomal HIV-1 infection intermediates in patients on highly active anti-retroviral therapy. Nat. Med. *6*, 76–81. 10.1038/71569.

131. Van Maele, B., De Rijck, J., De Clercq, E., and Debyser, Z. (2003). Impact of the Central Polypurine Tract on the Kinetics of Human Immunodeficiency Virus Type 1 Vector Transduction. J. Virol. *77*, 4685–4694. 10.1128/JVI.77.8.4685-4694.2003.

132. Omelina, E.S., Ivankin, A.V., Letiagina, A.E., and Pindyurin, A.V. (2019). Optimized PCR conditions minimizing the formation of chimeric DNA molecules from MPRA plasmid libraries. BMC Genomics *20*, 536. 10.1186/s12864-019-5847-2.

133. Liu, J., Song, H., Liu, D., Zuo, T., Lu, F., Zhuang, H., and Gao, F. (2014). Extensive Recombination Due to Heteroduplexes Generates Large Amounts of Artificial Gene Fragments during PCR. PLOS ONE *9*, e106658. 10.1371/journal.pone.0106658.

134. Farrell, A.G., Dadonaite, B., Greaney, A.J., Eguia, R., Loes, A.N., Franko, N.M., Logue, J., Carreño, J.M., Abbad, A., Chu, H.Y., et al. (2022). Receptor-Binding Domain (RBD) Antibodies Contribute More to SARS-CoV-2 Neutralization When Target Cells Express High Levels of ACE2. Viruses *14*, 2061. 10.3390/v14092061.

135. Liu, L., Wang, P., Nair, M.S., Yu, J., Rapp, M., Wang, Q., Luo, Y., Chan, J.F.-W., Sahi, V., Figueroa, A., et al. (2020). Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. Nature *584*, 450–456. 10.1038/s41586-020-2571-7.

136. Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., et al. (2021). Sustainable data analysis with Snakemake. 10.12688/f1000research.29032.2.

137. VanderPlas, J., Granger, B.E., Heer, J., Moritz, D., Wongsuphasawat, K., Satyanarayan, A., Lees, E., Timofeev, I., Welsh, B., and Sievert, S. (2018). Altair: Interactive Statistical Visualizations for Python. J. Open Source Softw. *3*, 1057. 10.21105/joss.01057.

138. Crawford, K.H.D., and Bloom, J.D. (2019). alignparse: A Python package for parsing complex features from high-throughput long-read sequencing. J. Open Source Softw. *4*, 1915. 10.21105/joss.01915.

139. Neher, R.A. (2022). Contributions of adaptation and purifying selection to SARS-CoV-2 evolution. Virus Evol. *8*, veac113. 10.1093/ve/veac113.

140. Turakhia, Y., Maio, N.D., Thornlow, B., Gozashti, L., Lanfear, R., Walker, C.R., Hinrichs, A.S., Fernandes, J.D., Borges, R., Slodkowicz, G., et al. (2020). Stability of SARS-CoV-2 phylogenies. PLOS Genet. *16*, e1009175. 10.1371/journal.pgen.1009175.

141. Simonich, C.A., Williams, K.L., Verkerke, H.P., Williams, J.A., Nduati, R., Lee, K.K., and Overbaugh, J. (2016). HIV-1 Neutralizing Antibodies with Limited Hypermutation from an Infant. Cell *166*, 77–87. 10.1016/j.cell.2016.05.055.

142. Umotoy, J., Bagaya, B.S., Joyce, C., Schiffner, T., Menis, S., Saye-Francisco, K.L., Biddle, T., Mohan, S., Vollbrecht, T., Kalyuzhniy, O., et al. (2019). Rapid and Focused Maturation of a VRC01-Class HIV Broadly Neutralizing Antibody Lineage Involves Both Binding and Accommodation of the N276-Glycan. Immunity *51*, 141-154.e6. 10.1016/j.immuni.2019.06.004.

143. Freund, N.T., Wang, H., Scharf, L., Nogueira, L., Horwitz, J.A., Bar-On, Y., Golijanin, J., Sievers, S.A., Sok, D., Cai, H., et al. (2017). Coexistence of potent HIV-1 broadly neutralizing antibodies and antibody-sensitive viruses in a viremic controller. Sci. Transl. Med. *9*, eaal2144. 10.1126/scitranslmed.aal2144.

144. Gao, F., Bonsignori, M., Liao, H.-X., Kumar, A., Xia, S.-M., Lu, X., Cai, F., Hwang, K.-K., Song, H., Zhou, T., et al. (2014). Cooperation of B Cell Lineages in Induction of HIV-1-Broadly Neutralizing Antibodies. Cell *158*, 481–491. 10.1016/j.cell.2014.06.022.

145. Doria-Rose, N.A., Altae-Tran, H.R., Roark, R.S., Schmidt, S.D., Sutton, M.S., Louder, M.K., Chuang, G.-Y., Bailer, R.T., Cortez, V., Kong, R., et al. (2017). Mapping Polyclonal HIV-1 Antibody Responses via Next-Generation Neutralization Fingerprinting. PLOS Pathog. *13*, e1006148. 10.1371/journal.ppat.1006148.

146. Georgiev, I.S., Doria-Rose, N.A., Zhou, T., Do Kwon, Y., Staupe, R.P., Moquin, S., Chuang, G.-Y., Louder, M.K., Schmidt, S.D., Altae-Tran, H.R., et al. (2013). Delineating Antibody Recognition in Polyclonal Sera from Patterns of HIV-1 Isolate Neutralization. Science *340*, 751–756. 10.1126/science.1233989.

147. Schommers, P., Gruell, H., Abernathy, M.E., Tran, M.-K., Dingens, A.S., Gristick, H.B., Barnes, C.O., Schoofs, T., Schlotz, M., Vanshylla, K., et al. (2020). Restriction of HIV-1 Escape by a Highly Broad and Potent Neutralizing Antibody. Cell *180*, 471-489.e22. 10.1016/j.cell.2020.01.010.

148. Caskey, M., Klein, F., and Nussenzweig, M.C. (2019). Broadly neutralizing anti-HIV-1 monoclonal antibodies in the clinic. Nat. Med. *25*, 547–553. 10.1038/s41591-019-0412-8.

149. Caskey, M. (2020). Broadly-Neutralizing Antibodies (bNAbs) for the Treatment and Prevention of HIV Infection. Curr. Opin. HIV AIDS *15*, 49–55. 10.1097/COH.0000000000000600.

150. Dadonaite, B., Crawford, K.H.D., Radford, C.E., Farrell, A.G., Yu, T.C., Hannon, W.W., Zhou, P., Andrabi, R., Burton, D.R., Liu, L., et al. (2022). A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike. 2022.10.13.512056. 10.1101/2022.10.13.512056.

151. Nduati, R., John, G., Mbori-Ngacha, D., Richardson, B., Overbaugh, J., Mwatha, A., Ndinya-Achola, J., Bwayo, J., Onyango, F.E., Hughes, J., et al. (2000). Effect of Breastfeeding and Formula Feeding on Transmission of HIV-1A Randomized Clinical Trial. JAMA *283*, 1167–1174. 10.1001/jama.283.9.1167.

152. Goo, L., Chohan, V., Nduati, R., and Overbaugh, J. (2014). Early development of broadly neutralizing antibodies in HIV-1–infected infants. Nat. Med. *20*, 655–658. 10.1038/nm.3565.

153. Simonich, C.A., Doepker, L., Ralph, D., Williams, J.A., Dhar, A., Yaffe, Z., Gentles, L., Small, C.T., Oliver, B., Vigdorovich, V., et al. (2019). Kappa chain maturation helps drive rapid development of an infant HIV-1 broadly neutralizing antibody lineage. Nat. Commun. *10*, 2190. 10.1038/s41467-019-09481-7.

154. Freed, E.O., and Martin, M.A. (1996). Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. J. Virol. *70*, 341–351. 10.1128/jvi.70.1.341-351.1996.

155. Tedbury, P.R., Ablan, S.D., and Freed, E.O. (2013). Global Rescue of Defects in HIV-1 Envelope Glycoprotein Incorporation: Implications for Matrix Structure. PLOS Pathog. *9*, e1003739. 10.1371/journal.ppat.1003739.

156. Scheid, J.F., Mouquet, H., Feldhahn, N., Seaman, M.S., Velinzon, K., Pietzsch, J., Ott, R.G., Anthony, R.M., Zebroski, H., Hurley, A., et al. (2009). Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. Nature *458*, 636–640. 10.1038/nature07930.

157. Bonsignori, M., Montefiori, D.C., Wu, X., Chen, X., Hwang, K.-K., Tsao, C.-Y., Kozink, D.M., Parks, R.J., Tomaras, G.D., Crump, J.A., et al. (2012). Two Distinct Broadly Neutralizing Antibody Specificities of Different Clonal Lineages in a Single HIV-1-Infected Donor: Implications for Vaccine Design. J. Virol. *86*, 4688–4692. 10.1128/JVI.07163-11.

158. Guo, H.H., Choe, J., and Loeb, L.A. (2004). Protein tolerance to random amino acid change. Proc. Natl. Acad. Sci. U. S. A. *101*, 9205–9210. 10.1073/pnas.0403255101.

159. Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adami, C., and Arnold, F.H. (2005). Thermodynamic prediction of protein neutrality. Proc. Natl. Acad. Sci. *102*, 606–611. 10.1073/pnas.0406744102.

160. Kuiken, C., Korber, B., and Shafer, R.W. (2003). HIV Sequence Databases. AIDS Rev. *5*, 52–61.

161. Effects of CCR5 and CD4 Cell Surface Concentrations on Infections by Macrophagetropic Isolates of Human Immunodeficiency Virus Type 1 | Journal of Virology https://journals.asm.org/doi/10.1128/JVI.72.4.2855-2864.1998?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed.

162. Derdeyn, C.A., Decker, J.M., Sfakianos, J.N., Wu, X., O'Brien, W.A., Ratner, L., Kappes, J.C., Shaw, G.M., and Hunter, E. (2000). Sensitivity of Human Immunodeficiency Virus Type 1 to the Fusion Inhibitor T-20 Is Modulated by Coreceptor Specificity Defined by the V3 Loop of gp120. J. Virol. *74*, 8358–8367. 10.1128/JVI.74.18.8358-8367.2000.

163. Wei, X., Decker, J.M., Liu, H., Zhang, Z., Arani, R.B., Kilby, J.M., Saag, M.S., Wu, X., Shaw, G.M., and Kappes, J.C. (2002). Emergence of Resistant Human Immunodeficiency Virus Type 1 in Patients Receiving Fusion Inhibitor (T-20) Monotherapy. Antimicrob. Agents Chemother. *46*, 1896–1905. 10.1128/AAC.46.6.1896-1905.2002.

164. Falkowska, E., Le, K.M., Ramos, A., Doores, K.J., Lee, J.H., Blattner, C., Ramirez, A., Derking, R., van Gils, M.J., Liang, C.-H., et al. (2014). Broadly Neutralizing HIV Antibodies Define a Glycan-Dependent Epitope on the Prefusion Conformation of gp41 on Cleaved Envelope Trimers. Immunity *40*, 657–668. 10.1016/j.immuni.2014.04.009.

165. Differential processing of HIV envelope glycans on the virus and soluble recombinant trimer | Nature Communications https://www.nature.com/articles/s41467-018-06121-4.

166. deCamp, A., Hraber, P., Bailer, R.T., Seaman, M.S., Ochsenbauer, C., Kappes, J., Gottardo, R., Edlefsen, P., Self, S., Tang, H., et al. (2014). Global Panel of HIV-1 Env Reference Strains for Standardized Assessments of Vaccine-Elicited Neutralizing Antibodies. J. Virol. *88*, 2489–2507. 10.1128/JVI.02853-13.

167. Landais, E., Huang, X., Havenar-Daughton, C., Murrell, B., Price, M.A., Wickramasinghe, L., Ramos, A., Bian, C.B., Simek, M., Allen, S., et al. (2016). Broadly Neutralizing Antibody Responses in a Large Longitudinal Sub-Saharan HIV Primary Infection Cohort. PLOS Pathog. *12*, e1005369. 10.1371/journal.ppat.1005369.

168. McCoy, L.E., and Burton, D.R. (2017). Identification and specificity of broadly neutralizing antibodies against HIV. Immunol. Rev. *275*, 11–20. 10.1111/imr.12484.

169. Scheid, J.F., Mouquet, H., Ueberheide, B., Diskin, R., Klein, F., Oliveira, T.Y.K., Pietzsch, J., Fenyo, D., Abadir, A., Velinzon, K., et al. (2011). Sequence and Structural Convergence of Broad and Potent HIV Antibodies That Mimic CD4 Binding. Science *333*, 1633–1637. 10.1126/science.1207227.

170. McGuire, A.T. (2019). Targeting broadly neutralizing antibody precursors: a naïve approach to vaccine design. Curr. Opin. HIV AIDS *14*, 294. 10.1097/COH.0000000000000548.

171. Derking, R., and Sanders, R.W. (2021). Structure-guided envelope trimer design in HIV-1 vaccine development: a narrative review. J. Int. AIDS Soc. *24*, e25797. 10.1002/jia2.25797.

172. Lee, J.H., Andrabi, R., Su, C.-Y., Yasmeen, A., Julien, J.-P., Kong, L., Wu, N.C., McBride, R., Sok, D., Pauthner, M., et al. (2017). A Broadly Neutralizing Antibody Targets the Dynamic HIV Envelope Trimer Apex via a Long, Rigidified, and Anionic β-Hairpin Structure. Immunity *46*, 690–702. 10.1016/j.immuni.2017.03.017.

173. Lynch, R.M., Wong, P., Tran, L., O'Dell, S., Nason, M.C., Li, Y., Wu, X., and Mascola, J.R. (2015). HIV-1 Fitness Cost Associated with Escape from the VRC01 Class of CD4 Binding Site Neutralizing Antibodies. J. Virol. *89*, 4201–4213. 10.1128/JVI.03608-14.

174. Leggat, D.J., Cohen, K.W., Willis, J.R., Fulp, W.J., deCamp, A.C., Kalyuzhniy, O., Cottrell, C.A., Menis, S., Finak, G., Ballweber-Fleming, L., et al. (2022). Vaccination induces HIV broadly neutralizing antibody precursors in humans. Science *378*, eadd6502. 10.1126/science.add6502.

175. da Silva, J., Coetzer, M., Nedellec, R., Pastore, C., and Mosier, D.E. (2010). Fitness Epistasis and Constraints on Adaptation in a Human Immunodeficiency Virus Type 1 Protein Region. Genetics *185*, 293–303. 10.1534/genetics.109.112458.

176. Wang, W., Nie, J., Prochnow, C., Truong, C., Jia, Z., Wang, S., Chen, X.S., and Wang, Y. (2013). A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. Retrovirology *10*, 14. 10.1186/1742-4690-10-14.

177. Chen, J., Kovacs, J.M., Peng, H., Rits-Volloch, S., Lu, J., Park, D., Zablowsky, E., Seaman, M.S., and Chen, B. (2015). Effect of the cytoplasmic domain on antigenic characteristics of HIV-1 envelope glycoprotein. Science *349*, 191–195. 10.1126/science.aaa9804.

178. Stano, A., Leaman, D.P., Kim, A.S., Zhang, L., Autin, L., Ingale, J., Gift, S.K., Truong, J., Wyatt, R.T., Olson, A.J., et al. (2017). Dense Array of Spikes on HIV-1 Virion Particles. J. Virol. *91*, e00415-17. 10.1128/JVI.00415-17.

179. Dingens, A.S., Arenz, D., Overbaugh, J., and Bloom, J.D. (2019). Massively Parallel Profiling of HIV-1 Resistance to the Fusion Inhibitor Enfuvirtide. Viruses *11*, 439. 10.3390/v11050439.

180. Pattnaik, G.P., and Chakraborty, H. (2020). Entry Inhibitors: Efficient Means to Block Viral Infection. J. Membr. Biol. *253*, 425–444. 10.1007/s00232-020-00136-z.

181. Mu, Z., Haynes, B.F., and Cain, D.W. (2021). Strategies for eliciting multiple lineages of broadly neutralizing antibodies to HIV by vaccination. Curr. Opin. Virol. *51*, 172–178. 10.1016/j.coviro.2021.09.015.

182. Dobson, C.S., Reich, A.N., Gaglione, S., Smith, B.E., Kim, E.J., Dong, J., Ronsard, L., Okonkwo, V., Lingwood, D., Dougan, M., et al. (2022). Antigen identification and high-throughput interaction mapping by reprogramming viral entry. Nat. Methods *19*, 449–460. 10.1038/s41592-022-01436-z.