Development and application of a phage-based deep mutational scanning system for mapping the fine epitopes of HIV-1 and SARS-CoV-2 specific antibodies

Meghan E. Garrett

A dissertation

submitted in the partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Julie Overbaugh, Chair

Erick Matsen

Justin Taylor

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Development and application of a phage-based deep mutational scanning system for mapping the fine epitopes of HIV-1 and SARS-CoV-2 specific antibodies

Meghan E. Garrett

Chair of the Supervisory Committee:

Affiliate Professor Julie Overbaugh

Professor, Divisions of Human Biology and Public Health Sciences, Fred Hutchinson Cancer Research Center

The rise of pandemic viruses such as HIV-1 and SARS-CoV-2 in modern times has highlighted the need for effective vaccines to combat these public health threats. In the case of HIV-1, development of an effective vaccine has been hampered by the virus' ability to rapidly evolve and evade detection by the immune system. While SARS-CoV-2 is not nearly as diverse a virus as HIV-1 and effective vaccines against SARS-CoV-2 virus do exist, early reports indicate that viral variants that escape vaccine-elicited immunity are emerging and may reduce the protective effect of vaccines. For both viruses there is a clear need to develop vaccines that elicit immunity that remains effective in the face of mutations.

Antibodies are a crucial component of immune response against HIV-1 and SARS-CoV-2. The protective effect of antibodies that develop after infection or immunization relies on their ability to either neutralize virus or direct the immune system to kill infected cells. Antibodies typically bind to the viral entry protein present on the surface of a virion or an infected cell, and mapping these binding sites is a key step needed to predict whether mutations on viral entry

proteins could lead to loss of antibody function. A myriad of methods can be used to map antibody epitopes, and each have various pros and cons. Structural methods such as X-ray crystallography or cryogenic electron microscopy (cryo-EM) are the gold standard for mapping antibody binding sites but are low-throughput and often slow. Other methods such as alanine scanning or peptide arrays do not give a complete picture of the effect of all amino acid mutations within antibody binding sites.

To accelerate mapping of antibody epitopes and potential escape mutations we developed Phage-DMS, a method of rapidly mapping the fine epitopes of monoclonal or polyclonal antibodies that combines phage display technology with deep mutational scanning (DMS). Phage-DMS involves the generation of a library of phage displaying peptides that correspond to either the wild-type sequence of a protein of interest or a sequence containing a single amino acid mutation. These libraries are incubated with antibody and then phage enriched by the antibody are sequenced, allowing us to determine epitope regions and sites sensitive to mutation.

In this thesis, I first describe the development and validation of Phage-DMS using four HIV-1 monoclonal antibodies that have been previously well characterized in the literature (Chapter 2). We compared the sites of escape as determined by Phage-DMS for these HIV-1 antibodies and found that our method recapitulated, and in some cases refined, the known epitopes of these antibodies that had been mapped using other methods. Additionally, we found that the effect of individual mutations as determined by Phage-DMS was a relatively quantitative measurement of the loss of binding as determined by a low throughput assay (peptide ELISA).

We then applied Phage-DMS to the study of antibodies in the plasma of people who had recovered from SARS-CoV-2 infection (Chapter 3). We found that convalescent antibodies commonly bound linear epitopes in the fusion peptide and heptad-repeat 2 regions of the Spike protein. Interestingly, examination of mutations that lead to escape from antibody binding revealed that there was person-to-person variation between the sites that were important for antibody binding. Additionally, we found that mutations within these conserved epitopes were not under selection in nature, suggesting either a lack of selective immune pressure or inability to evolve due to functional constraints at these sites.

Finally, we compared antibody epitopes and escape mutations in the plasma and sera of SARS-CoV-2 infected and mRNA vaccinated individuals (Chapter 4). We found that antibodies from vaccinated or severely infected individuals bound to linear epitopes in both the S1 and S2 subunits of Spike, whereas antibodies from mildly infected individuals bound to linear epitopes

in the S2 subunit alone. Antibody binding changed over time after vaccination within individuals, but was not affected by participant age, mRNA vaccine type, or vaccine dose. We also examined the effect of mutations within identified epitope regions and found that in many cases, vaccination induced a highly uniform escape profile across individuals. This finding has implications for the selection of escape variants on a global level.

In summary, the following chapters detail the creation and validation of Phage-DMS as a method of mapping antibody epitopes and escape mutations and describe its application with samples derived from SARS-CoV-2 infected and vaccinated individuals. The speed at which the COVID-19 pandemic has unfolded has underscored the need for methods that can provide answers equally as rapidly. With the capacity to screen hundreds of antibody samples in parallel and map epitopes down to the amino acid level in a short span of time, Phage-DMS is a significant step forward towards this goal. The studies in this thesis leverage this approach to define epitopes and pathways of escape for two of the greatest pandemic viruses in modern times.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I have many people to thank, all of whom supported me on my journey to graduate school and during my PhD studies.

A deep, deep thank you to my incredible mentor, Julie Overbaugh. I have matured into a better and more thoughtful scientist thanks to your patience, your keen advice, and your uncanny ability to both guide me and challenge me simultaneously. Thank you for putting your time and energy into improving every talk I've ever given and grant or paper I've ever written, your comments have always been invaluable. Your dedication to mentorship is inspiring, and I will carry the lessons I have learned in this lab throughout my life. Without exaggeration, coming into the Overbaugh lab as my 4th rotation was the best decision I ever made in grad school.

A major reason why my grad school experience has been so rewarding is due to the team of people Julie has assembled in the Overbaugh lab. Thank you to Dara Lehman for being a source of emotional support and useful advice at times when I was feeling low. A huge thank you to my past and present lab bay mates, Nicole Naiman and Zak Yaffe. Nicole mentored me during my rotation, patiently teaching me about the world of ADCC, and was a huge pillar of support during the first few years of my PhD. Zak has been such a fun person to share a lab bay with these last couple years, someone who always has a meme to make me laugh and who always is willing to let me bounce ideas off of him. And I need to apologize to both of them for always putting my feet up on their chair and getting in the way! Thank you to Ted Gobillot, who approached me during a poster session when he overheard me talking about potentially rotating in Julie's lab, and who probably changed the course of my scientific career. Thank you to my SARS-CoV-2 teammates, all of whom have been an enormous support during this pandemic: Caitlin Stoddard, Mackenzie Shipley, Hannah Itell, and Alex Wilcox. I've been lucky to work for two summers with my talented undergrad intern Margaret Chi, thanks for doing so much work on these projects. Thanks to Haidyn Weight and Vrasha Chohan for helping with the neutralization studies, they are just some of the many helpful technicians the lab has been fortunate enough to have. Thank you to Adam Dingens, for being an early mentor to me with the DMS project and for pushing me and my project forward. Even if I did not directly work with the other past and present Overbaugh lab members, they are all collectively the reason why I have been so happy to come into work every day: Stephanie Rainwater, Daryl Humes, Megan Stumpf, Caroline Kikawa, Laura Doepker, Cassie Simonich, Joshua Marceau, Sara Drescher, Sonja Danon, Caelan Radford, Jeremy Roop, Carolyn Fish, Bri Hennessey, Dana Arenz, Mark

Pankau, Feli Ruiz, Ryan Yucha. I have loved doing the NY Times crossword at lunch and dressing up for Halloween with all of you so much.

Thank you to the members of my committee, who have overseen the evolution of my projects over the years and who have focused my efforts: Jesse Bloom, Daphne Avgousti, Shiu-Lok Hu, and Justin Taylor. I would also like to thank Erick Matsen for joining my committee later and being another mentor for me with these Phage-DMS projects. Huge thanks to Jared Galloway and Kevin Sung in the Matsen group, who have been monumentally helpful by setting up computational pipelines to analyze the massive amounts of data we get from Phage-DMS experiments. Jared Galloway in particular was the co-author on the manuscript described in Chapter 4. And thanks to Kate Crawford, Allison Greaney, and Sarah Hilton in the Bloom lab for assistance with Phage-DMS and SARS-CoV-2 assays and samples over the years.

The research atmosphere and support at Fred Hutch has been so great, from Friday Beer Hour to the people in the Core facilities to the Thursday morning virus meeting to the Human Biology retreats and so much more. I've been lucky enough to serve on the Fred Hutch Student and Postdoc Advisory Committee, with guidance from Karen Peterson and Amber Ismael. The administrative support from the Fred Hutch graduate office along with the UW MCB graduate office has been so stellar thanks to Maia Low, Andrea Brocato, and many others. For our lab in particular, Shama Samant and Jasmine Gonzalez have been a huge help with everything from submitting grants to ordering to organizing meetings.

I would not be here today without the scientific mentors that I had before graduate school. Thank you to Craig Stephens at Santa Clara University, whose lab I worked in during a summer in high school and then again during my undergrad career. Thanks for being a kind and supportive mentor, and for giving me the advice I needed to make it to graduate school. Thank you to Rich James at Seattle Children's Research Institute, who took a chance and hired me as a research technician with no protein biology or cell culture experience and who taught me so much. I would not have been prepared for grad school without my time in the James lab.

I cannot write these acknowledgements without thanking my wonderful group of friends that have carried me throughout grad school. From my 611 roommates who created a home for me in Seattle and who put up with the MCB events I hosted, to my Monday night trivia teammates at Kate's Pub ("She Wants the PhD" forever!). You have all kept me sane and given me joy.

I feel intensely privileged to have the family that I do. Above all else, thank you to my parents, Leslie and Rick, for being my biggest cheerleaders since day 1, for believing in me and

supporting me in every way imaginable. Thank you to my sister, Cassidy, for keeping me grounded. Thank you to my grandma, Beverly O'Rourke, for taking me in when I moved to Seattle without a job and for being my home away from home here in Washington. And thank you to my grandpa, Jim O'Rourke, for instilling a love of learning in me and my entire family. I know he would be proud to see how far I've come. I have lots of family in Washington, including my aunts and uncles and cousin, and now my fiancé's family, who have all surrounded me with comfort and love.

To my fiancé, Kevin: thank you isn't enough. You have comforted me during those tough times, and you have celebrated my successes. Thank you for shoveling your car out of the snow and driving me to lab in in the middle of a snowstorm so I could finish that last experiment for my paper. Thank you for understanding when my experiments make me late to things. Thank you for listening to my rants about science and grad school, and I promise to listen to your rants about baseball and business school. My grad school journey wouldn't have been complete without finding you.

Chapter 1

INTRODUCTION

In the last 50 years, two viruses have caused devastating global pandemics: HIV-1 and SARS-CoV-2. Since its discovery in the 1980's, HIV-1 has been responsible for more than 30 million deaths worldwide [1]. Despite being the most-studied virus in history, an effective vaccine against HIV-1 remains elusive. In contrast, several vaccines were successfully developed and deployed against SARS-CoV-2 just one year after the first reported cases. Since the emergence of SARS-CoV-2, more than four million people have died in less than a two year period, and while the existence of effective vaccines has slowed the pace of this pandemic, the war is not yet over [2]. Emerging variants threaten to escape vaccine elicited immunity, and the search for a broadly protective coronavirus vaccine continues. In both cases, a deeper understanding of what a robust, protective immune response looks like is needed to gain control over these ongoing pandemics.

Below I introduce the biology of these two viruses, what is known about the antibody response to them, as well as current and novel methods used to study virus-specific antibodies.

1.1 HIV-1 and SARS-CoV-2 viral entry proteins

Viral entry proteins are proteins found on the surface of viruses that are responsible for mediating attachment to host receptors and/or entry into host cells. For enveloped viruses, such as HIV-1 and SARS-CoV-2, the viral entry protein is often the only virus-derived protein "visible" to the immune system on an intact virion and is thus a critical target of the humoral immune response. HIV-1 Envelope (Env) and SARS-CoV-2 Spike are both class I-type fusion glycoproteins responsible for attachment and entry for each respective virus. Class I fusion entry proteins generally consist of a trimeric assembly of heterodimers, with one subunit that mediates receptor attachment and one membrane-anchored subunit that mediates fusion [3]. Prior to fusion the membrane-anchored subunit must be proteolytically cleaved to expose the fusion peptide. The fusion peptide is made up of a hydrophobic stretch of amino acids that can insert into the host cell membrane after the viral entry protein attaches to the host receptor [4]. This then triggers an extension of the membrane-anchored subunit into a conformation known as the pre-hairpin intermediate state. This elongated structure subsequently collapses into a

lower energy hairpin conformation, bringing the host and viral membrane into proximity and driving fusion of the two membranes. At this final stage the membrane-anchored subunit forms a stable six-helical bundle.

Each viral entry protein faces unique evolutionary pressures to maintain these critical functions while also evading recognition by the immune system. Below I detail important structural information and functional capabilities of HIV-1 Env and SARS-CoV-2 Spike specific to each viral entry protein.

1.1.1 HIV-1 Envelope (Env)

The HIV-1 Env molecule is first synthesized as a precursor polypeptide, gp160, which is then proteolytically cleaved into two glycoproteins: gp120 and gp41 [5]. Gp120 is the surface exposed subunit responsible for receptor binding, and gp41 is the transmembrane subunit responsible for fusion. Gp120 and gp41 are decorated with N-linked glycans that are important for immune evasion, protein folding, and assist with binding to cells. After translation, glycosylation, and cleavage, mature Env is trafficked to the surface of the host cell where it is incorporated into a new budding virion. Upon binding of infectious virus to host receptors, Env will mediate fusion at the surface of the cell, and in addition Env is also capable of mediating cell-cell fusion.

HIV-1 Env specifically binds to the CD4 receptor as well as the CXCR4 or CCR5 co-receptors on the surface of human T cells and monocyte-derived cells, and the binding sites for these receptors are formed by tertiary folding of multiple domains on gp120, including conserved as well as more variable domains [6]. The variability of Env, and gp120 in particular, is a result of error-prone replication combined with selective pressure to escape the immune response [7]. HIV-1 employs several strategies to escape immune recognition, including but not limited to accumulation of mutations, addition of glycans, and lengthening or shortening of variable loop lengths on gp120 [8, 9].

Gp41 is a more conserved subunit containing the fusion machinery, and consists of an ectodomain, transmembrane domain, and the cytoplasmic domain [10]. The ectodomain contains the fusion peptide at the N-terminus along with two heptad repeat regions that drive the conformational changes required for fusion to occur.

1.1.2 SARS-CoV-2 Spike

The SARS-CoV-2 Spike glycoprotein is expressed and processed similarly to HIV Env, namely that it is produced first as a polypeptide and then is cleaved into two subunits that form a trimer of heterodimers: S1, which interacts with the host cell receptor, and S2, which contains the fusion machinery [11]. Two cleavage sites are present on Spike, the S1/S2 site and S2' site. Unlike other closely related coronaviruses, SARS-CoV-2 contains a furin cleavage motif at the S1/S2 boundary that allows the trimer to be "preactivated" while in the producer cell [12]. Subsequent cleavage at the S2' site is necessary for activation of the fusion peptide. Entry is thought to occur either at the surface of the cell, mediated by the serine protease TMPRSS2, or within endosomes, mediated by the lysosomal protease cathepsin. The main host receptor responsible for mediating entry is ACE2, although some studies have identified alternative host proteins that could bind to Spike and facilitate viral entry [13-16]. Therefore expression of these alternative viral entry proteins in cell lines versus primary cells could be an important consideration for assays with SARS-CoV-2 antibodies [17].

The S1 subunit of Spike contains 3 principal domains: the N-terminal domain (NTD), receptor binding domain (RBD), and C-terminal domain (CTD) [11]. The RBD contains the sites that directly interact with ACE2, whereas the functions of the NTD and CTD are not well understood. Most mutations to the Spike protein occur within the S1 protein, implying pressure by the immune system to escape recognition [18]. This variation may also reflect more functional tolerance for mutations in portions of the S1 protein as compared to S2.

Like HIV-1 gp41, the S2 subunit of Spike is a relatively more conserved protein that contains a fusion peptide along with two heptad repeat regions within the ectodomain. In order to stabilize the Spike protein for expression as a soluble protein, two proline mutations can be introduced to the S2 subunit in a region near the heptad repeat regions [19]. This first generation of stabilized Spike variants is known as S-2P, and these mutations are present in the Moderna mRNA Spike vaccine [20]. Another group has found that addition of four more proline mutations within the S2 subunit resulted in an even more stable and Spike molecule, which they named HexaPro [21]. These stabilized forms of Spike have accelerated vaccine design and have been useful for determining high resolution structures by electron microscopy.

1.2 Antibodies and immunity against HIV-1 and SARS-CoV-2

Antibodies, also known as immunoglobulin (Ig), are immune molecules made by B cells and consist of a pair of heavy chains and a pair of light chains. B cells that bind to antigen are

stimulated to replicate and iteratively improve binding in a process known as affinity maturation, resulting in antibodies that bind with high affinity to their target. After infection or immunization, antibodies against viral pathogens can typically be detected within two weeks. Antibodies can mediate various functions, depending on whether they bind to protein on the surface of the virion or to protein left on the surface of the cell after entry or before budding off as a new virion.

Eliciting antibodies is a main goal of vaccination and antibodies are also a major driver of antigenic drift in viral entry proteins. Below I give an overview of antibody functions and their role in protection against HIV-1 and SARS-CoV-2.

1.2.1 Antibody functions

Antibodies are “Y” shaped molecules. The tip of the “Y” is referred to as the Fab (fragment, antigen binding) region and contains variable sites that determine antigen specificity [22]. This is the portion of the antibody that binds to regions of HIV Env or SARS-CoV-2 Spike. The base of the “Y” is referred to as the Fc (fragment, crystallizable) region and interacts with immune cells and proteins by binding to an Fc receptor on the surface of cells. Antibodies can mediate a variety of biological functions through both the Fab and Fc regions, with a single molecule capable of mediating multiple functions. Antibody functions can be classified as either neutralizing or non-neutralizing:

Neutralizing functions

Neutralizing antibodies are antibodies that can block entry of a virion into a cell, preventing infection. The primary mechanism of neutralization is binding to the viral entry protein and physically blocking attachment to host cell receptors. Alternatively, some HIV-1 antibodies are known to be capable of inducing premature shedding of the gp120 protein from Env, thereby removing the key entry molecule from the virus surface [23]. Neutralization can also be achieved by binding to the fusion machinery and blocking the formation of the six helical bundle, as is done by some HIV-1 gp41-specific antibodies [24, 25].

Non-neutralizing functions

Non-neutralizing antibody functions rely on interactions between the antibody Fc region and immune effector cells or molecules [22]. The major mechanisms of action are through antibody dependent cellular cytotoxicity (ADCC), antibody dependent cellular phagocytosis (ADCP), and

antibody dependent complement deposition (ADCD). ADCC and ADCP can occur when antibody binds to antigen present on the surface of infected cells, whether left behind after entry or assembled at the plasma membrane before virus budding. For ADCC, interaction between the Fc region of the antibody and Fc receptors on immune cells such as neutrophils, monocytes, or macrophages activates release of cytotoxic granules and direct killing of infected cells. For ADCP, phagocytic cells are activated by binding to antibody Fc, triggering engulfment and destruction of the infected cell. ADCD is activated when C1q, a component of the complement system, binds to Ig-antigen complexes and triggers the complement cascade. ADCD can result in various outcomes including opsonization of the virion, deposition of membrane-attacking proteins, and/or recruitment of immune cells. The potency of non-neutralizing antibody functions is influenced by Ig isotype as well as modifications, such as glycosylation or fucosylation, of the Fc region [26]. Altogether, these non-neutralizing effector functions can help remove infected cells and in turn reduce viral levels during infection.

1.2.2 Antibody-mediated immunity to HIV-1

The first antibodies against HIV-1 Env to appear after infection are directed against gp41 and can be detected in plasma after about two weeks [27]. These early antibodies are non-neutralizing and do not appear to affect viral load. Non-neutralizing and weakly neutralizing antibodies directed against gp120 generally appear around a month after infection, but are quickly escaped by the evolving virus [28]. These early antibody responses also tend to be very specific to the infecting strain and do not recognize more diverse HIV-1 variants. In a subset of cases, individuals infected with HIV-1 can develop broadly neutralizing antibodies (bnAbs) capable of neutralizing diverse strains of HIV-1 [29]. BnAbs have been a major focus of HIV-1 vaccine efforts as it is thought that bnAbs are necessary to prevent establishment of HIV-1 infection.

There is some evidence that bnAbs could protect against infection, although data to support this in humans is limited. In macaque models, infusion with highly potent bnAbs has been shown to prevent acquisition of simian adapted HIV-1 (SHIV) [30-34]. However, results from the recent Antibody Mediated Protection (AMP) trial, where high-risk individuals were given infusions of a single bnAb targeting the CD4 binding site on Env, demonstrated that passive immunization only protected against acquisition of sensitive isolates of HIV-1 and did not provide protection more effectively than the placebo [35]. Correlates of protection have been

difficult to establish in humans as most HIV-1 vaccines have failed to show any efficacy, but evidence from one partially effective vaccine trial showed that only weakly neutralizing antibodies were elicited after immunization and were not likely responsible for the observed protection from infection [36].

In contrast, non-neutralizing antibody function has been shown to correlate with protection in many different contexts. In the same partially effective HIV-1 vaccine trial, infection risk was inversely correlated with high ADCC activity in the serum of vaccinees. Passive immunization studies have also implicated Fc mediated functions in protection from infection and death. Macaques infused with an antibody that mediates both neutralization and ADCC were protected from infection, but protection was reduced when ADCC functionality was specifically knocked out [37]. Infected mothers will passively transfer antibodies to their infants during pregnancy and breastfeeding, and HIV-specific ADCC activity has been shown to correlate with improved survival of infected infants [38, 39].

Overall, it is still unclear what kind of antibody response will be necessary in order to prevent acquisition of HIV-1, though any successful vaccine will likely need to elicit polyfunctional antibodies. To prevent escape, antibodies will also likely need to target multiple sites on HIV Env, ideally including responses to more functionally constrained regions. A better understanding of the activity and targets of protective antibodies is needed to achieve the goal of an effective HIV-1 vaccine.

1.2.3 Antibody-mediated immunity to SARS-CoV-2

Although the SARS-CoV-2 pandemic is much more recent in its origin than HIV-1, a wealth of information is already known about the antibody response against this virus. After infection, people develop antibodies against the Spike protein in a manner proportional to the severity of disease, with severely infected individuals producing high titers of binding and neutralizing antibodies [40, 41]. As expected, serum antibody titers wane over time after infection, but memory and plasma B cells have been found at 7 to 8 months after infection [42].

Antibodies, both neutralizing and non-neutralizing, have been found to be correlated with protection in several studies with animal models and humans. Transfer of convalescent serum into macaques significantly protected them from infection, and this correlated with neutralizing titer as well as non-neutralizing functions [43]. In hamster and mouse models, ablation of Fc mediated antibody activity reduced the protective capacity of human derived monoclonal

antibodies [44, 45]. Infection-derived immunity has been suggested to be protective, as a study of an outbreak on a fishing vessel demonstrated that those with preexisting neutralizing antibodies were completely protected from infection (non-neutralizing functions were not examined) [46]. Similarly, several studies of healthcare workers have found that prior infection reduced the likelihood of reinfection with SARS-CoV-2 [47, 48].

Vaccination with Spike mRNA induces high titers of binding antibodies, and in vaccinated individuals binding and neutralizing antibodies both correlated with protection [49, 50]. Interestingly, one study found that the emergence of non-neutralizing antibodies after vaccination was correlated with protection from infection during an early timepoint when vaccinated individuals did not have neutralizing antibodies [51]. Vaccine trials have also shown that even in the case of breakthrough infections, vaccinees were protected against severe disease.

Although antibody mediated immunity derived from infection or vaccination appears to be protective against current strains, this may not remain the case as variants of SARS-CoV-2 spread. Historical evidence from related endemic coronaviruses indicates that the Spike protein can evolve to escape antibodies elicited by prior infection, potentially explaining why endemic coronaviruses can reinfect the same host [52, 53]. The spread of the Delta variant of SARS-CoV-2 and the rise in breakthrough infections indicates that the excellent immunity induced by vaccination may not be robust in the face of a mutating virus.

1.3 Antibody epitopes

Given the important role antibodies play in guarding against infection with viral pathogens, there is a clear need to understand the interactions occurring at the interface between antibody and antigen. In modern vaccinology, antibodies with desired characteristics can act as a "blueprint" for an effective vaccine, a concept known as structure-based or reverse vaccine design [54]. In this system, information about antibody-antigen binding is used to design an antigen intended to elicit the prototypical antibody in a new host. Additionally, mapping the precise sites bound by pathogen-specific antibodies can aid in prospective surveillance of escape mutations.

The amino acids bound by an antibody are collectively known as the epitope and typically span 5-10 residues. Epitopes can be linear, whereby all epitope residues exist within a single continuous span on the protein, or discontinuous, whereby distal epitope residues are

brought together by the complex folding of protein chains. Below I discuss traditional methods of mapping antibody epitopes as well as a novel method called Phage-DMS which we have developed to identify key antibody binding sites as well as escape mutations.

1.3.1 Methods used to map antibody epitopes

There are many approaches used to map antibody epitopes, each with different pros and cons. The gold standard of epitope mapping is to image a structure of the antibody-antigen complex using X-ray crystallography or, more recently, cryogenic electron microscopy (cryo-EM) [55]. These methods allow you to see the orientation of the two molecules and predict the points of contact. While structure-based approaches can give atomic-level information, they are low-throughput and technically challenging. Furthermore, proximity between antibody and antigen side chains does not necessarily predict the "functional epitope", meaning the amino acids that would disrupt antibody function when mutated (this is further discussed in Chapter 2). Information on the functional epitope is needed to define mutations that may allow a virus to escape immune pressure.

Another common method of epitope mapping is peptide scanning, which employs a peptide microarray consisting of synthesized peptides immobilized onto a surface. As with other peptide-based methods, the major drawback is that this method can only identify linear epitopes. Additionally, it is not a particularly scalable approach and cannot achieve single-amino acid level detail of epitope specificity.

Mutagenesis of proteins or peptides is often used to identify single amino acids that contribute to antibody binding, and two traditional applications of this are site-directed mutagenesis and alanine scanning. Site-directed mutagenesis involves systematic replacement of single amino acids on a protein. Typically, however, this method is often driven by prior structural knowledge or studies reporting sites of viral escape. Alanine scanning is a version of site-directed mutagenesis whereby amino acids are replaced specifically with alanine, a small, chemically inert molecule that is less likely to disrupt overall protein structure. With this method, this process is repeated across the protein in order to identify residues important to some antibody function, such as neutralizing or binding. While these mutagenesis-based methods can yield detailed epitope maps, they are time and labor intensive and often can only be performed across a small region of the protein of interest.

1.3.2 Mapping antibody epitopes with deep mutational scanning

Deep mutational scanning (DMS) is a powerful method of mutagenesis that involves creating a library of all possible amino acid variants, subjecting the library to a selection, and then deeply sequencing the selected variants [56]. Our lab and others have harnessed DMS using several different systems to facilitate mapping of antibody epitopes.

Mutational antigenic profiling is a system where DMS libraries of viral entry proteins expressed by replication competent virus are passaged in the presence of neutralizing antibody and escape variants are detected. In this system, mutations that lead to loss of proper protein folding are purged from the library before passaging, thus resulting in identification exclusively of escape mutations that are tolerated by the virus. While mutational antigenic profiling has been employed to map HIV-1 and influenza specific neutralizing antibodies, it lacks the ability to map the functional epitope of non-neutralizing antibodies [57-60]. Other limitations of this method include the large amount of antibody required to perform mutational antigenic profiling, as well as the requirement to grow the library in biosafety level 2+ facilities.

Another application of DMS to epitope mapping has been performed using libraries of SARS-CoV-2 RBD variants expressed on the surface of yeast [61, 62]. Yeast with antibody bound to the displayed protein are sorted by flow cytometry and deeply sequenced. Like mutational antigenic profiling, only mutations to RBD that are tolerated and lead to protein expression on the surface of the yeast are profiled using this method. In this case, all binding antibodies can be mapped, although the yeast display system is low throughput due to the rate-limiting flow cytometry step. To date this method has only been used for portions of the Spike protein, namely the RBD, but not to study the full Spike protein.

To facilitate epitope mapping of all binding antibodies in a comprehensive and high throughput manner, we developed a method utilizing a DMS library of peptides displayed on the surface of phage called Phage-DMS. Peptides are first computationally designed to correspond either to the wild-type sequence or to sequences that contain a single amino acid mutation. These peptides are then synthesized on a DNA microarray and cloned into a phage display system. To perform a Phage-DMS experiment antibody is incubated with the library, immunoprecipitated, and then antibody-bound phage are deeply sequenced. Because phage can grow to high copy number, each binding reaction can take place in a single well of a 96 deep well plate, allowing potentially hundreds of antibody or serum samples to be screened in

parallel. The main drawback of this method is that due to the length of peptides displayed in Phage-DMS, only antibodies that include linear epitopes can be mapped.

We originally developed and validated this method in the context of HIV-1, as described in Chapter 2. We were then able to apply this method to characterize Spike-specific antibodies in samples from SARS-CoV-2 infected and vaccinated individuals (Chapters 3 and 4). By providing a high-throughput and comprehensive way of mapping antibody epitopes, we hope to accelerate antibody characterization and vaccine design. The rapid spread of SARS-CoV-2 has emphasized the importance of developing methods that can keep up with the pace of viral pandemics, and the studies described in this thesis provide an important step towards that goal. In this thesis I illustrate the application of this method for defining antibody escape pathways to these globally important viruses.

Chapter 2

PHAGE-DMS: A COMPREHENSIVE METHOD FOR FINE MAPPING OF ANTIBODY EPITOPES

Sections of text in this chapter have been modified slightly from the following manuscript:

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

2.1 Abstract

Understanding the antibody response is critical to developing vaccine and antibody-based therapies and has inspired the recent development of new methods to isolate antibodies. Methods to define the antibody-antigen interactions that determine specificity or allow escape have not kept pace. We developed Phage-DMS, a method that combines two powerful approaches – immunoprecipitation of phage peptide libraries and deep mutational scanning (DMS) – to enable high-throughput fine mapping of antibody epitopes. As an example, we designed sequences encoding all possible amino acid variants of HIV Envelope to create phage libraries. Using Phage-DMS, we identified sites of escape predicted using other approaches for four well characterized HIV monoclonal antibodies with known linear epitopes. In some cases, the results of Phage-DMS refined the epitope beyond what was determined in previous studies. This method has the potential to rapidly and comprehensively screen many antibodies in a single experiment to define sites essential for binding interactions.

2.2 Introduction

Antibodies are useful research tools, potential therapeutic molecules, and the end goal of many vaccines. Understanding the precise amino acids necessary for binding of antibody to its protein target can provide insights into pathways of escape and improve antigen design for vaccines. Defining these interactions can also enhance our knowledge of antibody function. Modern improvements in the isolation and cloning of monoclonal antibodies (mAbs) have resulted in a dramatic rise in the number of novel antibodies that can be produced, but current methods to map the epitopes of these antibodies cannot presently keep pace. Thus, there is a

need for a rapid screening tool to finely map the epitopes of many antibodies in a high-throughput manner.

Structural studies of antibody-antigen complexes are the gold standard for defining key amino acids that directly interact with an antibody, but typically are laborious and require large amounts of antibody. Recently, a method to display libraries of peptides on phage and probe for antibody binding via immunoprecipitation and deep sequencing has been described [63]. This method has been used to map the epitopes of novel HIV-specific mAbs [64-67], characterize the human virome [68], and discover autoantigens [69]. Phage libraries offer several advantages over peptide arrays and other mapping methods, namely that phage libraries are easy to generate and store, are relatively low cost, and can be used to rapidly screen for peptide-antibody binding with very small amounts of antibody or plasma. However, while these overlapping peptide libraries are useful for identifying an epitope region, they are limited in their ability to pinpoint individual residues critical for antibody binding.

Several methods exist to more precisely map the specific residues that define an antibody epitope, including the amino acids that disrupt binding and lead to immune escape. Alanine scanning gives single amino acid resolution of antibody epitopes, but it does not provide a complete picture of the potential effect of all possible amino acid mutations at a site. A more comprehensive way to understand the consequences of mutations within the epitope site is to use deep mutational scanning (DMS), which is a technique where each residue of a protein or peptide can be mutated to every possible variant [56]. The resulting library of variants is then used in a functional screen that simultaneously detects the impact of each mutation through deep sequencing. We have previously developed methods employing viral DMS libraries to map the epitopes of HIV-specific antibodies using neutralization as a functional screen [58]. While this approach detects viral escape from neutralization, it is not designed for use with antibodies that bind the viral antigen but mediate their effects through non-neutralizing functions. Therefore, creating a method of mapping antibody epitopes that measures binding agnostic of antibody function is needed.

We have built upon previous studies employing phage display in combination with DMS [70-73] and here we describe Phage-DMS, a new method that allows high-throughput and high-resolution mapping of antibody epitopes that are proximal in primary sequence with only a single round of immunoprecipitation. Using Phage-DMS, we identified the epitope of four well characterized HIV mAbs, confirming sites of escape predicted using other approaches as well

as finding novel epitope sites. Thus, Phage-DMS represents a new tool for the comprehensive mapping of antibody-antigen interactions.

2.3 Results

Generation and characterization of gp41/V3 HIV Envelope Phage-DMS libraries

A schematic summarizing Phage-DMS is depicted in Figure 2.1. In brief, we computationally designed sequences that varied at the central position so that this residue contained every possible amino acid corresponding to either the wild type residue or a mutant residue. These sequences were cloned into a phage display vector and then this phage display DMS library was incubated with the antibodies of interest. The resulting phage-antibody complexes were immunoprecipitated and the enriched phage were then subjected to deep sequencing to identify sequences specifically enriched or depleted in the presence of the antibody as compared to the initial phage display library.

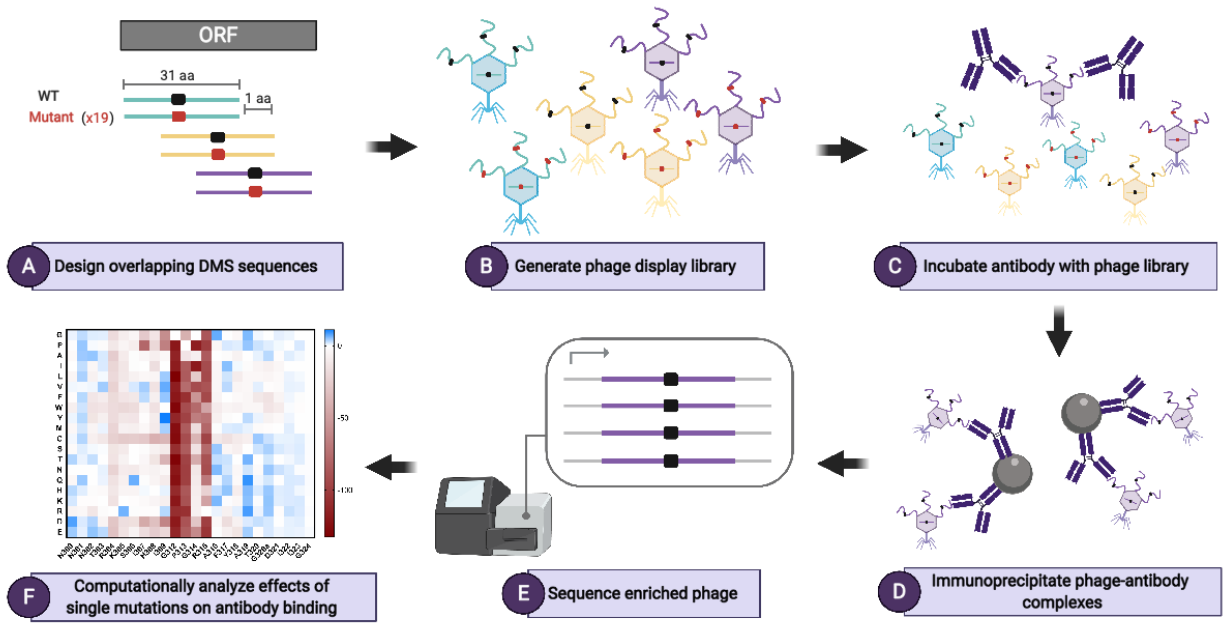


Figure 2.1. Schematic of Phage-DMS library design and experimental approach. (A) To build a Phage-DMS library, sequences are computationally designed to tile across an entire open reading frame of interest, with the central position varying to contain either the wild type residue (shown in black) or a mutant residue (shown in red). **(B)** Sequences are synthesized by releasable DNA microarray and then cloned into a T7 phage display vector. **(C)** The resulting phage display library is incubated with antibody, and **(D)** phage-antibody complexes are

immunoprecipitated with magnetic beads. **(E)** Sequences from enriched phage are PCR amplified and pooled samples are deeply sequenced. **(F)** Finally, computational analysis is performed to determine the relative effect of single mutations on the binding of antibody to antigen.

A Phage-DMS library encoding peptides from two immunodominant regions of HIV Env – the V3 region of gp120 and the gp41 ectodomain – was generated in the background of envelope sequences from two common HIV-1 circulating clades [74]: clade A (two variants: BG505, BF520) and clade C (ZA1197) (Supplemental Table 2.1). A total of 12,160 sequences were included. Duplicate libraries were made starting from the synthesized oligonucleotide array. Fifty-two percent of sequences from gp41/V3 library 1 and 37% of sequences from gp41/V3 library 2 matched the computationally generated sequences, based on sequencing individual plaques (Supplemental Figure 2.1A). The other plaque sequences contained either indels, point mutations, frameshifts, or multiple errors.

The representation of all expected unique sequences was determined by deep sequencing, performed at a depth of 82- and 45-fold coverage for gp41/V3 library 1 and 2, respectively. A high degree of all unique sequences (91.6% and 87.9%, respectively) were counted at least once in each library (Supplemental Figure 2.1B and C). There were some regions across gp41 and V3 that were not represented in the libraries, notably sequences corresponding to the fusion peptide domain of gp41. There were also differences in the representation of unique sequences by HIV strain, and we observed that peptides corresponding to BF520 Env were missing more than peptides from other strains. About half (51% and 50%, respectively) of all sequences detected were counted between 10 and 100 times within gp41/V3 library 1 and 2, indicating a modest degree of uniformity (Supplemental Figure 2.1D).

Mapping the epitope of gp41- and V3-targeting antibodies with Phage-DMS

To test the ability of our gp41/V3 Phage-DMS replicate libraries to define an antibody epitope, we performed experiments with Env-specific mAbs that have been previously characterized in detail and are known to bind to linear epitopes: gp41-specific mAbs 240D [75] and F240 [76]; V3-specific mAbs 447-52D [77] and 257D [78].

Gp41-specific mAbs

Deep sequencing of the phage bound by each of the gp41-specific antibodies showed enrichment of wild type sequences corresponding to the expected epitope region of gp41, the immunodominant C-C loop (Figure 2.2A and B). For both mAbs, peptides in the background of all three Env strains showed enrichment within the same region and to a similar degree. There was a high degree of correlation between technical replicates and good correlation between biological replicates, demonstrating the reproducibility of the Phage-DMS approach (Supplemental Figure 2.2).

To visualize the contribution of each amino acid to the epitope of the antibody, at each site we calculated the differential selection of each mutant-containing peptide as compared to the wild type-containing peptide, scaled by the strength of enrichment of the wild type peptide centered at that site, a metric hereafter termed “scaled differential selection”. Plotting the scaled differential selection for 240D and F240 in the context of the clade A Env protein (BG505) showed strong negative selection at positions C598 and C604 that form the disulfide C-C loop (Figure 2.2C and D). Mutations to either cysteine residue resulted in very low scaled differential selection values, with no amino acids showing higher preferred binding as compared to wild type, as expected for these C-C loop specific antibodies.

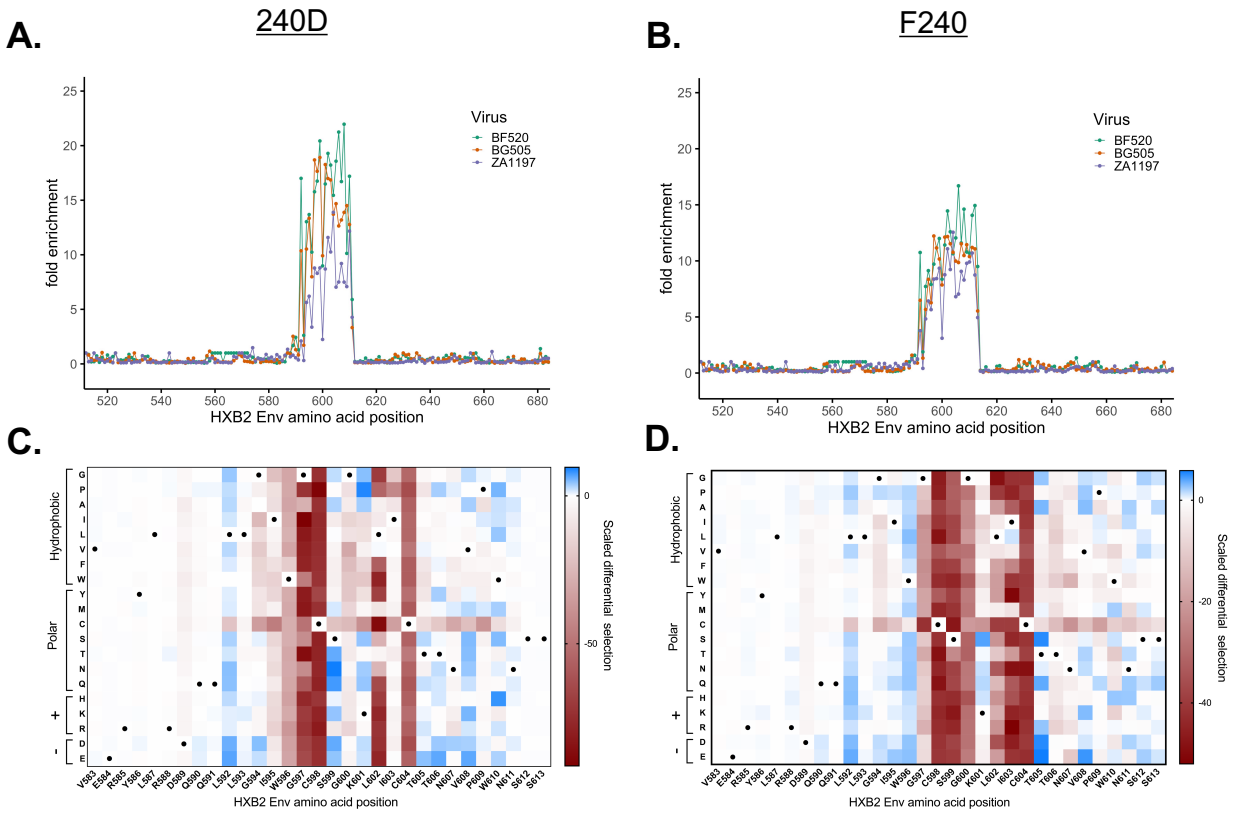


Figure 2.2. Enrichment and scaled differential selection results from Phage-DMS for gp41-specific mAbs with gp41/V3 libraries. **(A-B)** Line plot showing fold enrichment of wild type peptides in the background of each HIV Env strain for (A) mAb 240D and (B) mAb F240. The color corresponding to each wild type sequence is shown in the upper right. The x axis shows the amino acid position within HIV Env based on HXB2 reference sequence (C-D) Heatmap showing the relative effect, as compared to wild type BG505 Env, of each mutation on the binding to (C) mAb 240D and (D) mAb F240, within a selected region of gp41 shown with HXB2 numbering. Wild type residues are marked by a black dot. Amino acids are grouped based on their properties as indicated to the left. All data shown is the average of two biological replicates. See *Quantification and Statistical Analysis* for fold enrichment and scaled differential selection calculations.

For the 240D antibody, there was also selection against mutations to IWG (aa 595-597; Figure 2.2C), which was previously identified as the core epitope of 240D [75]. However, mutations to 602 were also negatively selected as compared to wild type, indicating that this site is also important for the epitope. There was enrichment above wild type for most mutations to position L592, as well as single amino acids that are preferred above wild type. For example, the variant N at position 599 and P at position 601 were preferred above wild type, illustrating the potential of this method to identify a sequence that presents a more optimal epitope sequence than the wild type viral sequence.

For the F240 antibody, nearly all mutations in C-C loop residues 597 through 604, with the exception of 601, resulted in depletion in Phage-DMS, suggesting they are part of the epitope (Figure 2.2D), consistent with structurally-defined interactions for F240 [79]. Additionally, some mutations in regions flanking the predicted epitope were positively enriched in Phage-DMS, for example sites 592 and 596.

V3-specific mAbs

Phage-DMS with the V3-specific mAbs 447-52D and 257D enriched for wild type peptides corresponding to the expected epitope region of the V3 loop [77, 80] (Figure 2.3A and B), with good correlation between replicates (Supplemental Fig 2.3). For 447-52D, there was binding only to wild type peptides from one of the clade A Env strains (BG505) and the clade C Env strain (ZA1197), but not the other clade A Env strain (BF520) (Figure 2.3A). This antibody has been reported to bind best to V3 sequences encoding the GPGR motif, but all the Env sequences in this library contain the less preferred GPGQ motif[81]. Sequences outside of this

core region differ between the strains that bind (IRIGPGQ) and the strain that does not bind (VHLGPGQ), suggesting these differences may influence binding to 447-52D in the context of the GPGQ motif. There was a similar result with 257D where only BG505 and ZA1197 V3 peptides within the expected epitope region were enriched, again indicating that naturally occurring sequence differences at the tip of the V3 loop may affect binding (Figure 2.3B).

The amino acid positions that showed the strongest negative selection with 447-52D in the context of BG505 were in the GPGQ core epitope (Figure 2.3C). The exception to this is at position 315, where R was the most strongly preferred amino acid. Additionally, variation at sites 304-309 outside of this motif was negatively selected, suggesting these amino acids contribute to the epitope in this context. In particular, mutations to 304 demonstrated consistent negative selection, which has not been reported.

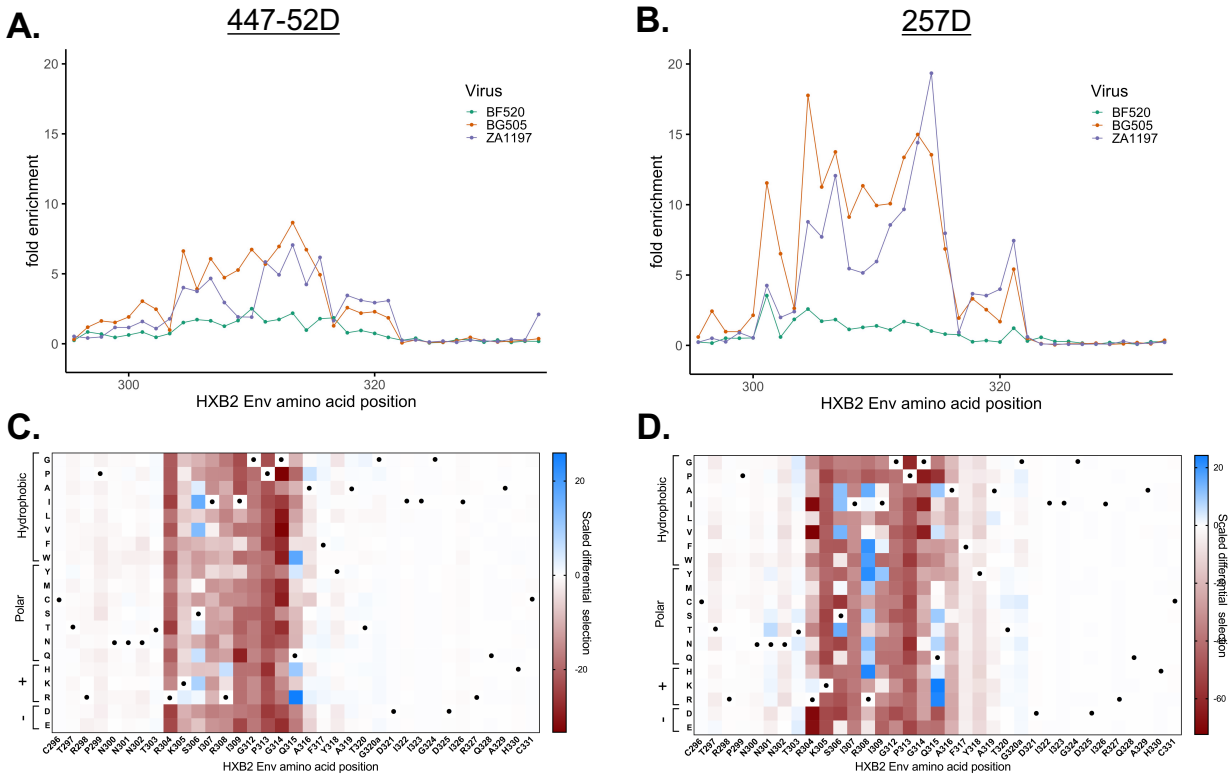


Figure 2.3. Enrichment and scaled differential selection results from Phage-DMS for V3-specific mAbs with gp41/V3 libraries. **(A-B)** Line plot showing fold enrichment of wild type peptides in the background of each HIV Env strain for (A) mAb 447-52D and (B) mAb 257D. The color corresponding to each wild type sequence is shown in the upper right. The x axis shows the amino acid position within HIV Env based on HXB2 reference sequence **(C-D)** Heatmap showing the relative effect, as compared to wild type BG505 Env, of each mutation on the binding to (C) mAb 447-52D and (D) mAb 257D, within the entire V3 region shown with HXB2 numbering. Wild type residues are marked by a black dot. Amino acids are grouped based on their properties as indicated to the left. All data shown is the average of two biological

replicates. See *Quantification and Statistical Analysis* for fold enrichment and scaled differential selection calculations.

Most mutations to residues spanning 304 to 316 showed strong negative selection with 257D in the context of BG505 (Figure 2.3D). There were also a number of mutations in this region that demonstrated binding above the wild type residue, including the R at position 315, consistent with the reported preference of 257D for the GPGR motif [82]. H at position 308 was also strongly enriched above the wild type sequence, which agrees with previous studies [80, 82]. Interestingly, there were several other amino acids at position 308 (e. g. F, Y, W) that were more enriched than the wild type R.

Generation of gp120 HIV Envelope Phage-DMS libraries and validation of V3 epitopes

To test whether we would find similar results with a different Phage-DMS library, a second replicate set of Phage-DMS libraries was constructed, this time containing all peptides spanning the gp120 domain of Env. Given that 9-13% of sequences were missing in the first library, we implemented two changes that were designed to minimize bias against sequences with higher GC content. First, we optimized the GC content of the sequences to be as uniform as possible, with an average GC content of 47%. Second, we implemented subcycling of the annealing and elongation steps between higher and lower temperatures during PCR because this was previously shown to significantly improve amplification of short template pools [83]. 28,840 sequences were generated in the background of a clade A (BG505), clade B (B41), and clade C (DU422) Env and cloned into duplicate phage libraries. Sequencing of individual plaques revealed that 47% of all sequences from gp120 library 1 and 41% of all sequences from gp120 library 2 were correct (Supplemental Figure 2.4A).

We determined the representation of sequences within each library by deep sequencing at a depth of 71- and 65-fold coverage, respectively. Almost all (96.4% and 96.5%, respectively) unique sequences were present in gp120 library 1 and 2 (Supplemental Figure 2.4B and C). 66.7% of sequences were counted between 10 and 100 times for gp120 Library 1 and 68.1% for gp120 library 2, demonstrating improved uniformity as compared to the gp41/V3 libraries, where approximately half were represented at this frequency (Supplemental Figure 2.4D). Our experiments also indicated that subcycling alone improved amplification (Supplemental Figure 2.4E).

Because the gp120 Phage-DMS libraries we generated contained V3 sequences, we used the same two V3-specific mAbs as above, which enabled the comparison of results between the gp120 libraries and the gp41/V3 libraries. Technical and biological replicates performed with the V3 mAbs and the gp120 Phage-DMS libraries had good correlation (Supplemental Figure 2.5). When measuring enrichment of wild type sequences for 447-52D, we observed only strong enrichment of wild type peptides from the clade B Env strain, which contains the preferred GPGR motif (aa 312-315) (Figure 2.4A). 447-52D did not enrich for V3 peptides from the two Env strains that encode GPGQ. For 257D, enrichment of wild type peptides showed binding of this antibody was highest for peptides from the clade B strain, but there was also weak binding to peptides from the clade A strain (Figure 2.4B). There was no binding of 257D to peptides from the clade C strain.

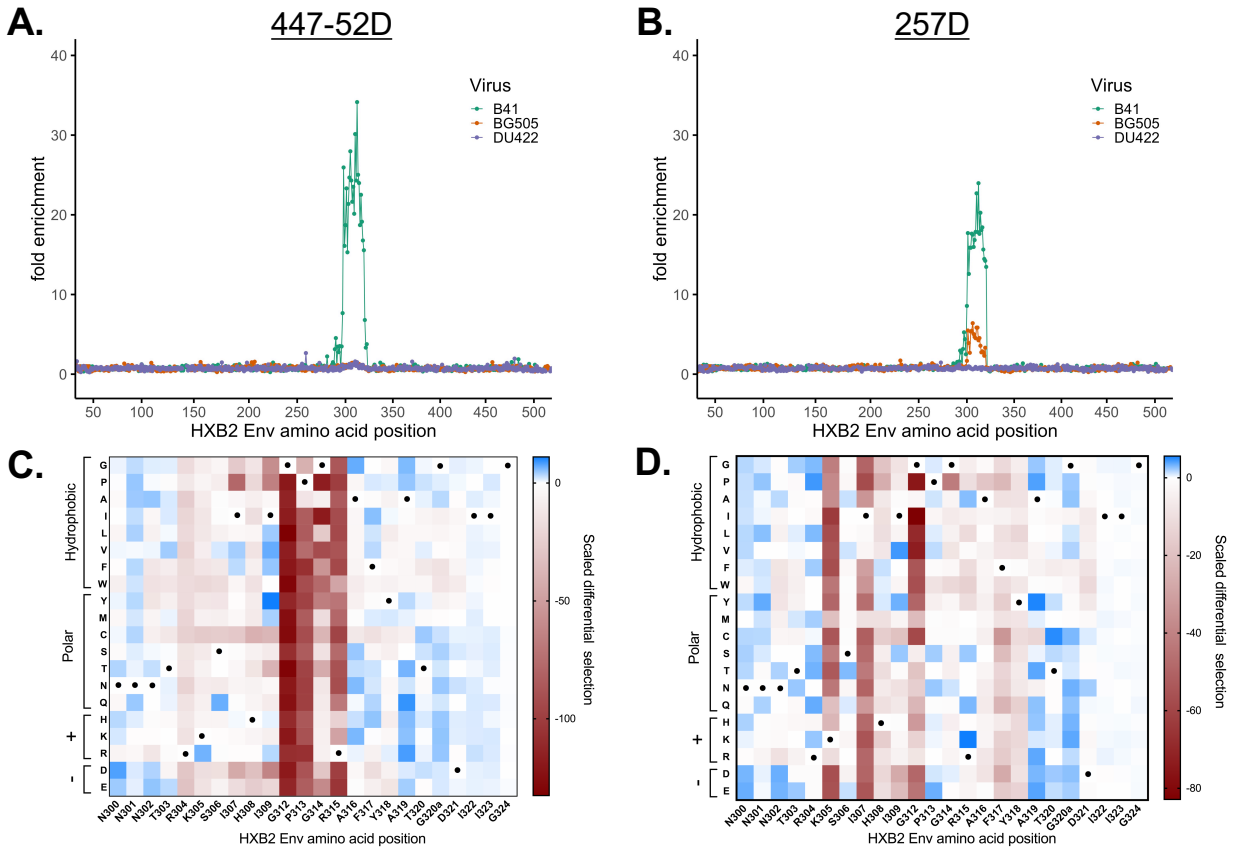


Figure 2.4. Enrichment and scaled differential selection results from Phage-DMS for V3-specific mAbs with gp120 libraries. **(A-B)** Line plot showing fold enrichment of wild type peptides in the background of each HIV Env strain for (A) mAb 447-52D and (B) mAb 257D. The color corresponding to each wild type sequence is shown in the upper right. The x axis shows the amino acid position within HIV Env based on HXB2 reference sequence (C-D) Heatmap showing the relative effect, as compared to wild type B41 Env, of each mutation on the binding to (C) mAb 447-52D and (D) mAb 257D, within a selected region of V3 shown with

HXB2 numbering. Wild type residues are marked by a black dot. Amino acids are grouped based on their properties as indicated to the left. All data shown is the average of two biological replicates. See *Quantification and Statistical Analysis* for fold enrichment and scaled differential selection calculations.

Mutations to the residues 312-315 at the GPGR tip resulted in strong negative selection in Phage-DMS experiments with 447-52D (Figure 2.4C). We additionally saw moderate negative selection at residues upstream of the GPGR, and in particular all mutations to site R304 and H308 resulted in negative selection. For 257D, strongest negative selection was observed for mutations at sites 305, 307, and 312 as compared to wild type (Figure 2.4D), which largely aligned with a previous study indicating that K-I--GP are the key residues for 257D binding (aa 305-313) [82]. The exception to this was P313, where we did not observe strong negative selection in the peptide context of this library, although we did with the gp120 library. We additionally saw possibly weak contribution of sites on the N-terminal side of the V3 loop (sites 317 and 318) to the epitope.

The observed epitope for both V3 antibodies with the gp120 Phage-DMS libraries was, in general, consistent with previous results with the gp41/V3 libraries. For 447-52D, negative selection at sites 312-315 agreed with previous findings from the gp41/V3 libraries, though the relative strength of negative selection differed between the libraries. For 257D, fewer sites within the epitope region were sensitive to mutation in the context of clade B Env in the gp120 library when compared to results with the gp41/V3 library in the background of clade A Env (BG505). For example, we did not observe that mutations to P313 reduced binding, as we did with the gp41/V3 library, indicating its contribution to the epitope may depend on the context.

Validation of Phage-DMS results using mutant peptides in a competition ELISA

To confirm that mutations identified by Phage-DMS disrupt binding to the V3- and gp41-specific antibodies, we performed competition enzyme-linked immunosorbent assays (ELISAs) with wild type and mutant V3 or gp41 peptides (see Supplemental Table 2.2). In this assay, peptides that bind to the antibody block activity against a coated antigen (gp120 or gp41); peptides with mutations in key epitope residues will no longer bind the antibody resulting in an increased IC50 value as compared to a wild type peptide, and peptides that have improved binding to antibody will result in a decreased IC50 value as compared to a wild type peptide.

The wild type gp41 peptide competed for binding of both antibodies to gp41 protein (Supplemental Figure 2.6A and B). The Phage-DMS results indicated strong preference for the

cysteines that form the loop structure for both antibodies, and indeed a gp41 peptide with a mutation in one of these cysteines (C598D) no longer competes for binding to the antibodies at any concentration tested (Figure 2.5A). Site L602 was also identified as part of the epitope based on a strong depletion of peptides containing mutations to that site. For both 240D and F240, a peptide with a mutation at this site exhibited a 103- and 22- fold increase in IC50, respectively. The IC50 value of peptide S599K was similar to the wild type peptide for 240D but was not detectable for F240, suggesting that the mutant peptide had has lost is ability to bind F240, consistent with the differences observed for these two antibodies by Phage-DMS. A negative control gp41-binding antibody (5F3) did not show reduced activity in the presence of any gp41 peptides (Supplemental Fig 2.6C).

We confirmed that the wild type V3 peptide competed for binding with both antibodies to gp120 protein (Supplemental Figure 2.6D and E). With both 447-52D and 257D, the R315W peptide containing a mutation within the important GPGR motif resulted in a 72- and 5-fold increase in IC50, respectively (Figure 2.5B). Interestingly, the peptide containing a mutation at site 304, which was identified by Phage-DMS as being part of the epitope for mAb 447-52D but not found by any previous studies, demonstrated a 21-fold increase in IC50 as compared to the wild type peptide. Sites 317 and 318 were identified by Phage-DMS as possibly weakly contributing to the epitope, and peptides containing mutations in those sites had a small but statistically significant difference in IC50 from the WT. Peptides with mutations in sites that were not identified by Phage-DMS as being part of the epitope, such as K305T with 447-52D and R304E with 257D, did not yield IC50 values that were significantly different from the wild type peptide in competition ELISAs. VRC01 was used as a negative control gp120-binding antibody and did not show reduced binding in the presence of any V3 peptides (Supplemental Figure 2.6F).

In order to examine whether the sites predicted to disrupt binding by Phage-DMS correlate with the binding results with corresponding peptides by ELISA, we compared IC50 values obtained for each WT and mutant peptide and differential selection values as measured by Phage-DMS. Differential selection measured for 240D and F240 in the context of the gp41/V3 library and for 447-52D and 257D in the context of the gp120 library each correlated well with the IC50 values measured by competition peptide ELISA (Figure 2.5C and D), with R^2 values of 0.74 and 0.85 respectively ($p=0.013$ and $p=4.2e-06$).

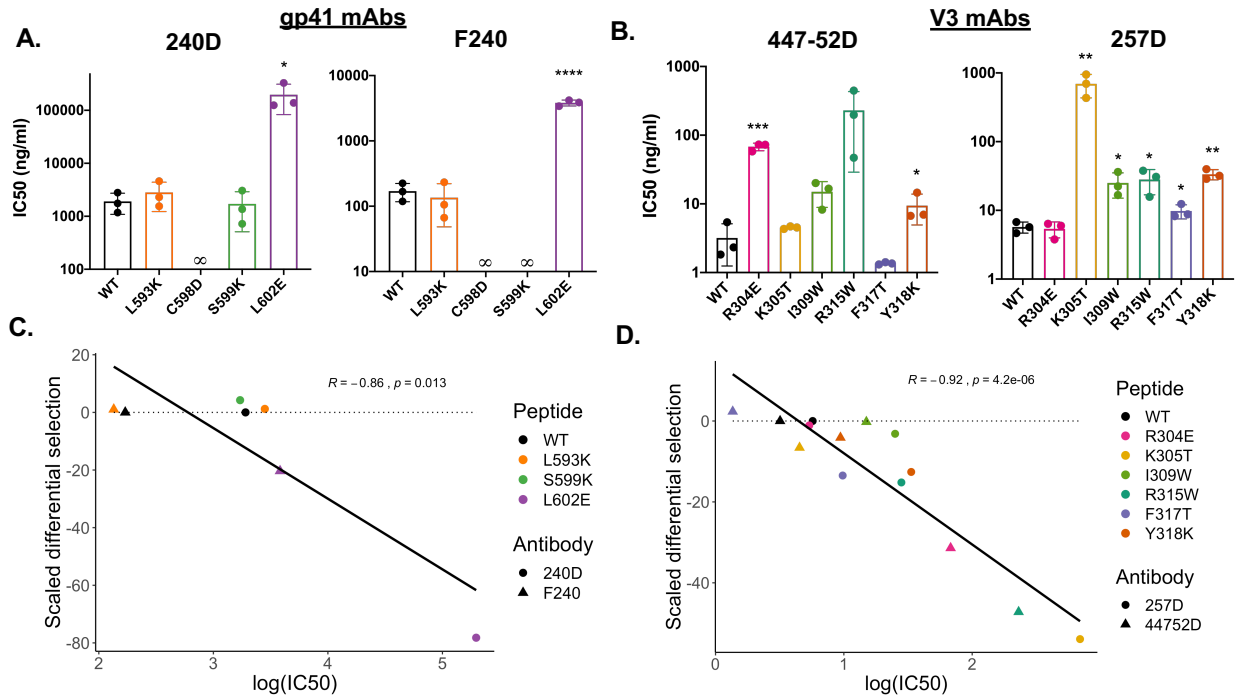


Figure 2.5. Comparison of effects of mutations on peptide binding in competition ELISAs and Phage-DMS. (A-B) Bar plot showing the IC50 values of wild type and mutant peptides in a competition ELISA for (A) gp41-specific mAbs with wells coated with MN gp41 protein or (B) V3-specific mAbs with wells coated with SF162 gp120 protein. Antibodies were pre-incubated with each peptide before addition to the wells. Peptides for which no inhibition of antibody binding could be detected at any concentration tested are indicated with an infinity symbol. Results for three replicate experiments are shown, with the mean +/- SEM. Statistical significance was determined by an unpaired t test. **(C-D)** Correlation between results of IC50 values from competition peptide ELISAs and scaled differential selection values from Phage-DMS for (C) gp41-specific mAbs and (D) V3-specific mAbs. Pearson's correlation shown at the top.

2.4 Discussion

The process of mapping the epitope of antibodies tends to be labor-intensive with current methods, and often does not provide a detailed understanding of the epitope. To address this, we have combined phage display technology and deep mutational scanning to create an approach called Phage-DMS that is capable of providing a comprehensive view of an antibody epitope in a high-throughput manner. As proof-of-concept, we created Phage-DMS libraries that display wild type and mutant peptides derived from HIV Env strains, and we used these libraries to define the individual amino acids critical for binding to HIV-specific antibodies. The effects of specific mutations identified in this way were validated using the more traditional approach of peptide binding by ELISA. The binding measurements were highly correlated with

Phage-DMS scaled differential selection, showing that Phage-DMS measurements capture mutational effects in a relatively quantitative way.

In addition to accurately identifying epitope residues that have been characterized by low-throughput mutagenesis and gold standard structural methods, Phage-DMS also refined and expanded upon the findings from previous studies. For example, mutations to R304 consistently resulted in decreased binding to 447-52D as compared to the wild type sequence, yet this site had not been described to comprise the epitope in previous studies using overlapping hexapeptides [77] or X-ray crystallography [81, 84]. While neither crystal structure showed electron density at site R304, the prediction from Phage-DMS was validated by competition peptide ELISA, suggesting that Phage-DMS captures information other than physical proximity between antigen and antibody. Conversely, there were instances where previous studies had defined sites on Env with close proximity and hydrogen bonding to the antibody by X-ray crystallography, such as residues 595 to 609 for F240 [79], but we did not observe strong negative selection taking place at sites 595, 596, or 605 - 609. Other studies have demonstrated that side chain proximity does not necessarily predict the contribution of a site to the formation of a protein-protein complex [57, 85-91], and this case with F240 further illustrates how analysis by Phage-DMS can yield information on the critical binding interactions of an antibody distinct from the identity of interfacing sites obtained by structural studies.

Identification of an antibody's epitope by Phage-DMS was reproducible, as demonstrated by the consistent results obtained by testing the same antibodies with two independently generated libraries. However, we did observe differences in the results with libraries built with different viral strains, highlighting the sensitivity of this assay to viral context. This is not surprising given that the strains included in the library differed by as much as 28% in their envelope sequences and thus the amino acid introduced by DMS will be influenced by these other sequences within the epitope. An example of this is 447-52D, where one library had a viral strain with a more optimal core epitope (B41 in the gp120 library), but the other library (gp41/V3) did not. In this case, other sites sensitive to mutation were revealed in the gp41/V3 library, presumably because none of the wild type sequences contained the optimal core motif. 257D is another potential example of the importance of strain context and/or assay in determining fine antibody epitopes. We observed that P313 contributed to binding with 257D in the gp41/V3 library, but not with the gp120 library. P313 had previously been identified as a key part of the epitope of 257D in a study using biopanning with random phage-displayed peptides [82], but was not recognized as contributing to binding in a different study that used overlapping

hexapeptides from the clade B strain, MN [80]. The variable importance of site P313 could be related to HIV strain and/or to differences between methods or assay sensitivity.

Phage-DMS has several advantages over other mapping approaches. The high titer and renewable nature of phage libraries make them ideal for large-scale and high-throughput experiments. While other groups have combined yeast display and deep mutational scanning to create a powerful tool for defining protein interactions [92, 93], these methods do have some limitations compared to Phage-DMS for these purposes. One major difference in throughput is because yeast display requires sorting by flow cytometry, which is time consuming and expensive, limiting the number of antibodies that can be tested, as previously noted [93]. By contrast, Phage-DMS experiments can be conducted in a 96 well plate format with an independent test of an antibody in each well. In addition, there is often a requirement for long-read sequencing for yeast display approaches, which increases costs and complexity of analysis compared to Phage-DMS, which relies on short sequence reads. Yeast also have lower transformation efficiency and typically yield 10^7 to 10^9 transformants per μg of DNA [94]. This creates a lower threshold for the number of variants that can be tested in a library as compared to phage, which can yield ~ 10^{11} transformants, more than 100-fold higher than yeast. Therefore, with Phage-DMS, antibodies can be easily tested against multiple background sequences in parallel, which is especially important when studying antibodies against a pathogen as diverse as HIV. Other mutation-based approaches often involve creating and testing mutants in the background of a single Env strain. Phage-DMS can identify the binding sites for neutralizing antibodies, like 447-52D and 257D, and non-neutralizing antibodies, like 240D and F240, whereas many other methods usually rely on a functional read out, such as the ability to escape neutralization. Moreover, unlike methods that involve growth of the virus in culture, phage will display peptides encoding mutations irrespective of their effect on replication fitness, resulting in a more complete picture of the potential effect of all variants on binding. Given its versatility, Phage-DMS is not limited to mapping antibody epitopes and could theoretically be used to map binding sites between any two proteins of interest, similar to other studies that have utilized DMS to dissect protein-protein interactions using phage [70-73].

There are some limitations of Phage-DMS. For one, post-translational modifications such as glycosylation are not present on the peptides displayed by the phage. Additionally, due to the length of peptides displayed in these libraries, Phage-DMS in its current form is limited to mapping epitopes that dependent on proximal amino acids and thus is not suited for complex conformational epitopes. To confirm these limitations, we tested known glycan-dependent

(BF520.1, QA013.2) and discontinuous epitope (PGT145, 50-69) HIV Env-specific mAbs in Phage-DMS experiments, and as expected did not observe enrichment of any peptide sequences (data not shown). However, our data and findings by others who have generated phage display libraries suggest that peptides can fold and present native-like conformations to some extent [69]. For example, Phage-DMS showed a strong preference for cysteines that form a predicted loop structure within the epitope region of the gp41-specific antibodies.

Phage-DMS can be applied not just towards identifying epitope sites and potential pathways of antibody escape, but also to improving the binding of antibody to antigen. We observed cases where mutations, such as F317T with antibody 447-52D, improved binding above wild type levels by both Phage-DMS and competition peptide ELISA, demonstrating that Phage-DMS results could optimize antigen targeting and aid in rational vaccine design. Overall, we have demonstrated that Phage-DMS can map sites essential for antibody binding and could potentially accelerate the characterization of mAbs and development of vaccines.

Limitations of the study

Because the Phage-DMS libraries described here display short linear peptides of 31 amino acids in length, generally this method is not optimal for identifying the full epitope of antibodies that rely on conformational presentation of the antigen. The peptides displayed in a Phage-DMS library will not undergo post-translational modification and therefore antibodies that rely on binding to glycans and other similar moieties cannot be defined using this method.

2.4 Methods

Antibody production

The following antibodies were obtained from the AIDS Reagent Program, Division of AIDS, NIAID: 5F3, F240, 240D, VRC01, 257D, 447-52D.

Generation of Phage-DMS libraries

HIV-1 strains included in the gp41/V3 DMS phage library and gp120 DMS phage library are noted in Supplemental Table 2.1. We did not include the signal peptide of gp120 and the transmembrane and cytoplasmic domains of gp41 in these libraries. Oligonucleotides coding for peptides 31 amino acids in length and containing a single variable residue were computationally designed to overlap by 30 amino acids and tile across the entire length of these protein

sequences. Sequence coding for a linker ($[G_4S]_3$ for gp120 and gp41 sequences, and 15 flanking wild type residues for the V3 sequences) was added to the beginning and end of the protein sequence of interest in order to ensure that every residue of the protein was located at the central position of a peptide [95]. All sequences were codon optimized for expression in *E. coli* by either using IDT online Codon Optimization Tool (http://www.idtdna.com/CodonOpt) for the gp41/V3 libraries or the Codon Optimization OnLine tool (COOL, http://cool.syncti.org/) for the gp120 libraries. Linker and protein coding sequences for the gp120 DMS phage library were additionally GC optimized using the COOL tool, with a target GC concentration of 50%. The central codon of each oligo was replaced with codons representing every possible amino acid, generating a total of 20 oligos centered at each site along the protein. Each oligo additionally had 5' and 3' adaptor sequences added to facilitate amplification and cloning (5': AGGAATTCTACGCTGAGT, 3': TGATAGCAAGCTTGCC). After de-duplicating oligonucleotides that were identical across the different HIV strains, 12,160 unique sequences for the gp41/V3 library and 28,840 unique sequences for the gp120 library were synthesized as an oligonucleotide pool by Twist Bioscience. Code used to generate oligonucleotide sequences can be found at github.com/meghangarrett/Phage-DMS.

Oligonucleotide pools were resuspended in Tris-EDTA (TE) buffer, pH 8.0, to a concentration of 10 ng/uL. PCR amplification of the pools with the KAPA HiFi PCR Kit (Roche) was done with 2.5 uL of a 2 ng/uL solution of the pool in a 50uL reaction, with primers annealing to the adaptor sequences (Fwd: AATGATACGGCAGGAATTCCGCTGAGT, Rvs: CGATCAGCAGAGGCAAGCTTGCTATCA). Amplification of the gp41/V3 oligos was performed using the following thermocycler conditions:

1. 95°C for 3 min
2. 98°C for 20 s
3. 60°C for 15 s
4. 72°C for 15 s
5. Go to step 2 (x3)
6. 95°C for 30 s
7. 98°C for 20 s
8. 72°C for 30 s
9. Go to step 7 (x20)
10. 72°C for 5 min
11. Hold at 4°C

Amplification of the gp120 library was performed including subcycling steps (adapted from [83] , using the following thermocycler conditions:

1. 95°C for 3 min
2. 98°C for 20 s
3. 60°C for 15 s
4. 72°C for 15 s
5. Go to step 2 (x3)
6. 95°C for 30 s
7. 98°C for 20 s
8. 67°C for 15 s
9. 72°C for 15 s
10. Go to step 8 (x4)
11. Go to step 7 (x20)
12. 72°C for 5 min
13. Hold at 4°C

PCR products were cleaned using Agencourt AMPure XP beads (Beckman Coulter) and then digested with EcoRI-HF and HindIII-HF (NEB) at 37°C for 1 hour. Cloning into bacteriophage was performed using the T7Select System (EMD Millipore). In brief, four separate digests of the PCR products were performed and then pooled, purified using gel electrophoresis, and ligated at a 3:1 vector to insert ratio into the T7Select 10-3b bacteriophage vector arms overnight at 16°C. 5 uL of the ligation reaction was added 25 uL of T7Select Packaging Extract and incubated at room temperature for 2 hours, then stopped using 270 uL sterile LB. The phage titer from the packaging reaction was determined using a plaque assay, then amplified on the host E coli (BLT5403) and titered again. At every step each library member was represented by >200 plaque forming units (pfu) to avoid bottlenecking, and the final libraries had >200,000 pfu/mL per unique library member. Sanger sequencing of at least 40 plaques from each library was performed to assess the diversity and fidelity of the phage libraries. This process was performed in duplicate, starting from the PCR amplification step, and the final libraries were stored at -80°C with 0.1 volumes sterile 80% glycerol and penicillin/streptomycin added.

Antibody immunoprecipitation with Phage-DMS library

Immunoprecipitation with antibody bound to peptide displayed on phage was performed in a manner previously described [63, 64]. In brief, the phage library was first thawed and diluted to a concentration representing 200,000 pfu/mL per unique library member (for example, the gp120 DMS phage library has 28,840 members, therefore we use a concentration of 5.8×10^9

pfu/mL). After blocking a deep 96 well plate with 3% BSA in TBST, 1 mL of the diluted phage library with 10 ng of the antibody of interest was added to the well. Plates were sealed and incubated on a rocker at 4°C for 20 hours. 20 uL of each Protein A and Protein G Dynabeads (ThermoFisher) were added to each well and incubated on a rocker at 4°C for 4 hours. Magnetic separation was performed and antibody-bead complexes were washed 3x with 400 uL wash buffer (150 mM NaCl, 50 mM Tris-HCl, 0.1% [vol/vol] NP-40, pH 7.5). Beads were resuspended in 40 uL of water, transferred to a 96 well PCR plate and then bound phage lysed at 65°C for 10 minutes. Antibody selections were done in technical duplicate, and duplicate mock selected samples were included in order to determine the background levels of peptide binding to the beads. Additionally, we lysed 10-20 million phage from the diluted input library to determine the distribution of phage in the starting library.

Deep sequencing

To determine the frequency of each peptide in the antibody selected, mock selected, and input conditions, we deep sequenced the lysed phage from each sample. We performed two rounds of PCR to amplify and add dual barcodes to each sequence. Each PCR reaction was performed using Q5 Hot Start High Fidelity 2X Master Mix (NEB). For the first round of PCR, 20uL of lysed phage was used as the template in a 50 uL reaction along with primers that anneal to sequences on either side of the cloning region within the T7 Select 10-3b vector (Fwd: TCGTCGGCAGCGTCTCCAGTCAGGTGTGATGCTC, Rvs: GTGGGCTCGGAGATGTGTATAAGAGACAGCAAGACCCGTTTAGAGGCCC). We used the following thermocycler conditions for the first round of PCR:

1. 98°C for 30 s
2. 98°C for 5 s
3. 68°C for 10 s
4. 72°C for 30 s
5. **Go to step 2 (x30)**
6. 72°C for 2 min
7. Hold at 4°C

2 uL of the PCR product was then used as the template in a 50 uL reaction for the second round of PCR, with indexed P5 and P7 primers that anneal to sequence added by the round 1 primers (Fwd: AATGATACGGCGACCACCGAGATCTACACNNNNNNNNTCGTCGGCAGCGTCTCCAGTC,

Rvs:
CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTCTCGTGGGCTCGGAGATGTGTATAAG
AGACAG, where “NNNNNNNN” denotes a unique 7-nt indexing sequence). We used the following thermocycler conditions for the second round of PCR:

1. 98°C for 30 s
2. 98°C for 5 s
3. 72°C for 40 s
4. Go to step 2 (x8)
5. 72°C for 2 min
6. Hold at 4°C

PCR products were then cleaned using Agencourt AMPure XP beads and eluted in 50 uL water. DNA concentrations were quantified via Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher). Equal amounts of DNA from the antibody selected samples were pooled, along with equal amounts of mock selected and input library at 10X the amount of the antibody selected samples (for example, 20 ng of each antibody selected sample was pooled along with 200 ng of each mock selected sample and input library sample). Finally, the pooled sample was gel purified, quantified using the KAPA Library Quantification Kit (Roche), and then sequenced on an Illumina MiSeq with 1x125 bp single end reads using a primer that anneals to the 5’ adaptor sequence just prior to the oligo sequence (GCTCGGGGATCCGAATTCTACGCTGAGT).

Competition peptide ELISA

Custom peptides were synthesized by ThermoFisher. V3 peptides were resuspended in water at 1 mg/mL, and gp41 peptides were resuspended in 50% DMF due to increased hydrophobicity and to prevent degradation of the cysteine disulfide bond. To perform a competition peptide ELISA, plates were first coated overnight with 500 ng/mL of either SF162 gp120 (Cambridge Biologics, # 01-01-1063) or MN gp41 (AIDS Reagent Program, # 12027) in PBS. After washing 4X with wash buffer (1X PBS, 0.05% Tween-20), wells were blocked with blocking buffer (1% BSA in PBS + 1mM EDTA) for 1 hour at 37°C. Antibodies diluted to 500 ng/mL in blocking buffer were preincubated for 30 minutes with peptide at various concentrations, with controls incubated with equivalent volume of appropriate peptide resuspension buffer to establish baseline activity. Blocking solution was washed from plates and then antibody-peptide samples were added to each well and the plate was incubated for 1 hour

at 37°C. Plates were washed and then goat anti-human HRP (Sigma) diluted 1:2500 in blocking buffer was added to each well and incubated for 1 hour at 37°C. After washing again, 1-Step Ultra TMB-ELISA Substrate Solution (ThermoFisher) was added and incubated at room temperature for either 15 minutes (V3 peptides) or 5 minutes (gp41 peptides) before adding an equivalent volume of 1N H_2SO_4. OD_{450} value for each well was then measured on a plate reader. All assays were performed in 384 well plates (ThermoFisher) and washed by robot. IC50 values are calculated by taking the OD_{450} values across a dose response curve of the peptide and computing in Prism (GraphPad version 8.4.0) the concentration of peptide which inhibits half of the maximum activity (e.g. activity with no peptide present).

Quantification and Statistical Analyses

Code used to perform the following analyses and generate plots with Phage-DMS data can be found at https://github.com/meghangarrett/Phage-DMS.

Demultiplexing and alignment

Demulitplexing and fastq file generation were performed using Illumina MiSeq Reporter software. Reads were aligned to their respective reference libraries using bowtie v1.1.1 [96], with the options "--trim3 32 -n 0 -l 93 --tryhard --nomaqround --norc --best --sam –quiet". We aimed to get 10X coverage of antibody selected samples and 100X coverage of mock selected samples and input library, keeping in mind that only about half of the phage in the library contained sequences that perfectly matched the computationally designed sequences. All reads were deposited to NCBI and are accessible in the Sequence Read Archive: PRJNA630833.

Calculating enrichment of peptides

The enrichment ($E_{r,x}$) of a peptide containing amino acid x centered at site r is calculated as follows. First, read counts n from technical replicates are added together. The proportion p of each peptide within a sample (Equation 1) or within the input library (Equation 2) is calculated by comparing the read count of each peptide with amino acid x at site r to the sum of all read counts in the sample.

$$p_{r,x}^{selected} = (n_{r,x}^{selected} + f^{selected} \times P)/(\sum_y n_{r,y}^{selected} + f^{selected} \times P) \quad (1)$$

$$p_{r,x}^{input} = (n_{r,x}^{input} + f^{input} \times P)/(\sum_y n_{r,y}^{input} + f^{input} \times P)$$

(2)

To account for statistical noise, a pseudocount of $P = 1$ is added to each count. To scale the pseudocount according to the varying sequencing depth of the antibody selected sample and library input, we calculated $f^{selected}$ (Equation 3) and f^{input} (Equation 4) variables as follows.

$$f^{selected} = max\left[1, (\sum_y n_{r,y}^{selected})/(\sum_y n_{r,y}^{input})\right]$$

(3)

$$f^{input} = max\left[1, (\sum_y n_{r,y}^{input})/(\sum_y n_{r,y}^{selected})\right]$$

(4)

To calculate the enrichment $E_{r,x}$, the proportion $p_{r,x}^{selected}$ of each peptide within the antibody selected sample is compared against the proportion $p_{r,x}^{input}$ of each peptide within the input library (Equation 5)

$$E_{r,x} = p_{r,x}^{selected}/p_{r,x}^{input}$$

(5)

Calculating differential selection and scaled differential selection

Differential selection is a metric used to calculate the relative enrichment of a peptide containing a mutation ($E_{r,x}$) as compared to the enrichment of a peptide containing the wild type amino acid at a site ($E_{r,wt(r)}$) (Equation 6).

$$d_{r,x} = \log_2(E_{r,x}/E_{r,wt(r)})$$

(6)

By definition, the differential selection of the wild type amino acid at a site ($d_{r,wt(r)}$) is always 0.

In order to emphasize the differential selection taking place at sites within the epitope, we calculated a metric we termed “scaled differential selection”. To get the scaled differential selection $S_{r,x}$ of a mutation x at a site of interest r, we take the differential selection as calculated above and multiply it by the enrichment of the peptide containing the wild type amino acid centered at the position of interest (Equation 7).

$$S_{r,x} = d_{r,x} \times E_{r,wt(r)} \quad (7)$$

Graphical illustrations

Graphical abstract and Figure 2.1 were made with BioRender.com.

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Julie Overbaugh (joverbau@fredhutch.org).

Materials Availability

All unique reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

Data and Code Availability

Sequencing data has been deposited to NCBI and are accessible under BioProject # PRJNA630833. Aligned sequencing data and all code used to perform analyses generate plots with Phage-DMS data can be found at https://github.com/meghangarrett/Phage-DMS.

2.5 Notes

Acknowledgments

We thank T. Gobillot and A. Dingens for helpful discussions, C. Sather and the Fred Hutch Genomics Core for assistance with sequencing, and L. Stamatatos for generous use of lab

equipment. This work was supported by NIH R01 AI138709 and AI120961. MEG was supported in part by T32 AI083203.

Contributions

JO conceived the project; MEG and JO led the design of the study; MEG and HLI contributed to experimental design and performed experiments; MEG, KDC, RB and JDB participated in computational and data analyses. MEG and JO wrote the paper with input from all authors.

Declaration of Interests

MEG and JO are inventors on a patent application on Phage-DMS.

Chapter 3

HIGH RESOLUTON PROFILING OF PATHWAYS OF ESCAPE FOR SARS-COV-2 SPIKE-BINDING ANTIBODIES

Sections of text in this chapter have been modified slightly from the following manuscript:

Garrett, M.E., Galloway, J., Chu, H.Y., Itell, H.L., Stoddard, C.I., Wolf, C.R., Logue, J.K., McDonald, D., Weight, H., Matsen IV, F.A., Overbaugh, J. High-resolution profiling of pathways of escape for SARS-CoV-2 spike-binding antibodies. *Cell* **184**, 2927-2938.e2911, doi:10.1016/j.cell.2021.04.045 (2021).

3.1 Abstract

Defining long-term protective immunity to SARS-CoV-2 is one of the most pressing questions of our time and will require a detailed understanding of potential ways this virus can evolve to escape immune protection. Immune protection will most likely be mediated by antibodies that bind to the viral entry protein, Spike (S). Here we used Phage-DMS, an approach that comprehensively interrogates the effect of all possible mutations on binding to a protein of interest, to define the profile of antibody escape to the SARS-CoV-2 S protein using COVID-19 convalescent plasma. Antibody binding was common in two regions: the fusion peptide and linker region upstream of the heptad repeat region 2. However, escape mutations were variable within these immunodominant regions. There was also individual variation in less commonly targeted epitopes. This study provides a granular view of potential antibody escape pathways and suggests there will be individual variation in antibody-mediated virus evolution.

3.1 Introduction

The global outbreak of a novel coronavirus, SARS-CoV-2, has claimed over one million lives within just one year after the first detected case (https://coronavirus.jhu.edu/map.html), with countless others experiencing long-term health problems after recovering from infection. Vaccines to prevent SARS-CoV-2 spread and treatments to reduce disease severity are currently under rapid development, with many strategies relying on antibody-mediated immunity. The main viral target of interest for vaccines and antibody therapies against SARS-CoV-2 is the

coronavirus spike (S) protein, which decorates the surface of the virion and mediates attachment and entry into host cells [97]. The S protein is comprised of a trimeric assembly of two subunits: S1 and S2, which are proteolytically cleaved at the S1/S2 boundary. The S1 subunit contains an N-terminal domain (NTD) and a receptor binding domain (RBD) within the C-terminal domain (CTD). The S2 subunit contains the fusion peptide (FP) along with two heptad repeat regions (HR1 and HR2), separated by a linker region, responsible for driving viral and host membrane fusion [98, 99]. Binding of the S1 protein via the RBD to the human ACE2 receptor is followed by proteolytic cleavage at the S2' site, which exposes the FP and activates a series of conformational changes resulting in membrane fusion [12, 100, 101].

Neutralizing antibodies targeting the SARS-CoV-2 RBD have been the main focus of vaccine strategies and antibody therapies, as they block virus entry in cell culture [102-105] and prevent infection or disease in some animal models [106-108]. However, the study of other coronaviruses has illustrated that antibodies elicited by infection can target epitope regions outside of the RBD. For example, a number of neutralizing antibodies directed to regions other than the RBD have been isolated from individuals infected with the closely related viruses SARS-CoV and MERS-CoV [109]. Recent SARS-CoV-2 studies of serum antibodies from COVID-19 patients have led to the identification of neutralization activity directed at linear epitopes found just downstream of the RBD, overlapping the FP, and just upstream of HR2 in the linker region [110, 111]. Thus, there may be multiple regions within the SARS-CoV-2 S protein that may shape the viral immune response.

While antibodies to the RBD are a logical initial focus for studies of protective antibodies against SARS-CoV-2, it is not yet known whether they are a correlate of protection for SARS-CoV-2 in humans. A limited understanding of protective immunity is to be expected at the early stage in this new disease, and a broad view is therefore prudent. In this regard, it is important to note that studies from other viruses such as HIV and Ebola have shown that non-neutralizing antibodies are an immune correlate of protection in humans [36, 38, 112, 113]. For SARS-CoV-2, antibodies against the S protein likely perform functions other than neutralization given that there is not a direct correlation between the levels of binding to S protein and neutralization titers [114]. Given the incomplete picture we have regarding immunity against SARS-CoV-2, it is important to study the antibody response irrespective of function and/or epitope in order to fill these knowledge gaps.

Given the rapid spread and amplification of SARS-CoV-2 in the population and the high mutation rate of RNA viruses, variants that can evade the immune response are likely to arise.

There is strong evidence that immune selection drove the emergence of escape mutants in the S protein of SARS-CoV [115] and MERS-CoV [116], raising the possibility that the same could occur with SARS-CoV-2. Immune selection may be further enhanced by a vaccine if it is not fully protective, making understanding the potential escape pathways of the virus critically important.

There are several recent studies that have examined the effect of select mutations on serum antibody binding to SARS-CoV-2. One study that assessed 82 S protein variants present in circulating SARS-CoV-2 detected a few mutations that resulted in decreased neutralization sensitivity [117]. However, only a small fraction of the potential mutations that could arise on the S protein were tested. Another study harnessed the power of deep mutational scanning (DMS) to capture a complete picture of the functional consequences of single mutations within the RBD on protein expression, ACE2 binding, and monoclonal antibody binding [118]. However, no study has yet examined the effect of the polyclonal antibody response on immune escape across the S protein, which is the target in current prophylactic and therapeutic strategies to combat COVID-19.

Previously we developed a comprehensive method of mapping escape mutations for HIV monoclonal antibodies, referred to as Phage-DMS [119]. In Phage-DMS, a library of all possible mutations to a protein is generated in peptide fragments, which are expressed by phage. Complexes of antibodies that bind the phage library are immunoprecipitated and sequenced to determine the antibody binding region(s) and the mutations within that epitope region that disrupt binding. Phage-DMS has several advantages: it is high-throughput, allowing comparison amongst a large number of samples in parallel, it can determine pathways of escape for both neutralizing and non-neutralizing antibodies that bind to linear epitopes, and results using this method also correlate well with mutational effects measured in other assays [119]. Here, we used Phage-DMS to understand the spectrum of single mutations on the S protein that could reduce antibody binding and thus mediate escape from plasma antibodies found in COVID-19 patients. Using a Phage-DMS library displaying both wildtype and mutant peptides tiling across the S protein, we identified a spectrum of single mutants that were capable of reducing antibody binding and found person-to-person variability in the effect of mutations within immunodominant epitopes. In the arms race between the humoral immune response and SARS-CoV-2, these results allow us to predict pathways of escape and forecast the appearance of escape mutants.

3.3 Results

Generation of the Spike Phage-DMS library

In order to explore all amino acids that define the epitope of antibodies directed towards the spike (S) protein of SARS-CoV-2, we generated a Phage-DMS library designed to tile across the ectodomain of the S protein from the Wuhan Hu-1 strain (Figure 3.1). We also included DMS peptides generated in the context of the D614G mutation, as clinical and in vitro evidence suggests that this variant may have increased infectivity as compared to the original Wuhan Hu-1 strain [120]. We computationally designed sequences coding for peptides 31 amino acids long with the variable amino acid in the central position. To achieve single amino acid resolution of epitope boundaries, peptides were designed to overlap by 30 amino acids. 24,820 unique peptides were designed in total; the peptide library included wildtype peptides that could be used to define the antibody epitope and peptides with all possible mutations to determine those within the defined epitope that disrupt or enhance antibody binding. Two biological replicate libraries of these peptide sequences were cloned as we have done previously [119]. Deep sequencing of the final duplicate libraries (Library 1 and Library 2) indicated that each contained a high percentage of all unique sequences (96.0% and 95.9%, respectively) (Supplemental Figure 3.1).

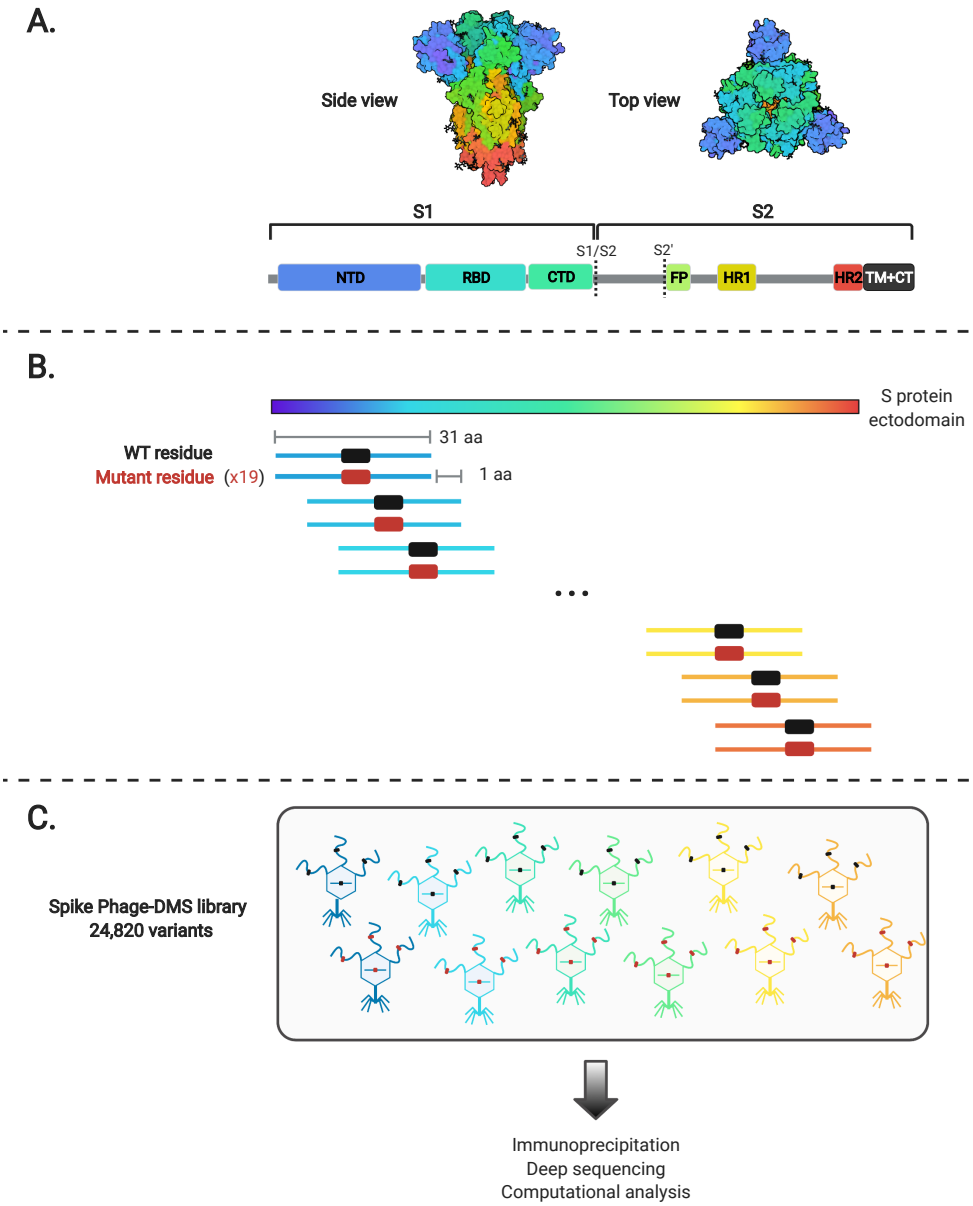


Figure 3.1. Schematic of the design of the Spike Phage-DMS library. (A) Structure of the S protein and location of important protein domains. Structure was made in BioRender.com (PDB: 6VXX). (B) Sequences were computationally designed to code for peptides 31 amino acids long and to tile stepwise across the Wuhan-Hu-1 SARS-CoV-2 S protein ectodomain by 1 amino acid. There are 20 peptides representing all 20 possible amino acids at the central position, containing either the wild type residue (shown in black) or a mutant residue (shown in red). Within the 31 aa region surrounding the D614G mutation, peptides were also generated with G614 in addition to the 20 amino acid variants at the central position. (C) The designed sequences were cloned into a T7 phage display vector and amplified to create the final S protein Phage-DMS library. This library was then used in downstream immunoprecipitation and deep sequencing experiments with human plasma. See Supplemental Figure 3.1 for distribution of sequences in the final library.

Enrichment of immunodominant linear epitopes by antibodies from COVID patients

We used the Spike Phage-DMS library to determine the pattern of antibody binding in plasma from a cohort of 18 COVID-19 patients from the HAARVI study in the Seattle area collected between March and May 2020 (Supplemental Table 3.1). Most patients had mild symptoms not requiring hospitalization, with the exception of one patient (6) who had moderate symptoms and required supplemental oxygen. Convalescent plasma was collected twice, at approximately day 30 and day 60 post symptom onset (p.s.o.; Supplemental Table 3.1).

To define the epitope region targeted by antibodies, we examined the enrichment of wildtype peptides from the library. We first calculated Pearson's correlation coefficient for the peptide enrichment values from each biological replicate experiment. In general, samples with poor correlation between replicate experiments were those that lacked reproducibly strong antibody binding. The correlation ranged from 0.96 to 0.27, with three patients having no sample from either timepoint above a correlation of 0.5 (Supplemental Figure 3.2A). The latter three cases (7, 16, and 17) were excluded from further analyses. Paired samples taken from the same individual at 30 and 60 days p.s.o. had peptide enrichment values that were well correlated, with a median correlation of 0.87, and were significantly better correlated than randomly paired samples ($p = 1.2e-09$, Wilcoxon rank sum test), which had a median correlation of 0.29 (Supplemental Figure 3.2B). The correlation between paired samples also tended to be stronger when the correlation between biological replicates, and therefore antibody binding, was stronger (Supplemental Figure 3.2C).

Wildtype peptides from two immunodominant regions, which include the FP and upstream sequences spanning aa 809-834 as well as the linker region and N-terminal portion of the adjacent HR2 domain spanning aa 1140-1168, were the most enriched peptides (Figure 3.2). Peptide enrichment data is also available in interactive form using the dms-view online tool [121] at https://github.com/meghangarrett/Spike-Phage-DMS/tree/master/analysis-and-plotting/dms-view. Interestingly, while most patients showed enrichment primarily within the linker region upstream of HR2, patient 5 showed strong enrichment of peptides from an epitope within the HR2 domain itself. The enriched peptides included aa 1167-1191 and this epitope was unique compared to the other 14 patients. We also observed cases where patient plasma enriched peptides from a less common epitope, albeit weakly, at both 30 and 60 days p.s.o.. Plasma from patient 12 enriched peptides within a region of the NTD (aa 255-280), patient 15

enriched peptides within both the RBD (aa 485-500) and the region just downstream of the RBD (aa 540-573), and patient 3 enriched peptides in the region just upstream of the S1/S2 cleavage site (aa 620-644) (Figure 3.2). Because most patients did not have antibodies that bound linear peptides within the RBD, despite it being a known target of antibodies, we confirmed by ELISA that the patients studied here indeed have plasma antibodies that target RBD (Supplemental Figure 3.3A). This indicates that antibodies targeting RBD bind to mainly conformational and/or glycosylated epitopes, which are missed by Phage-DMS.

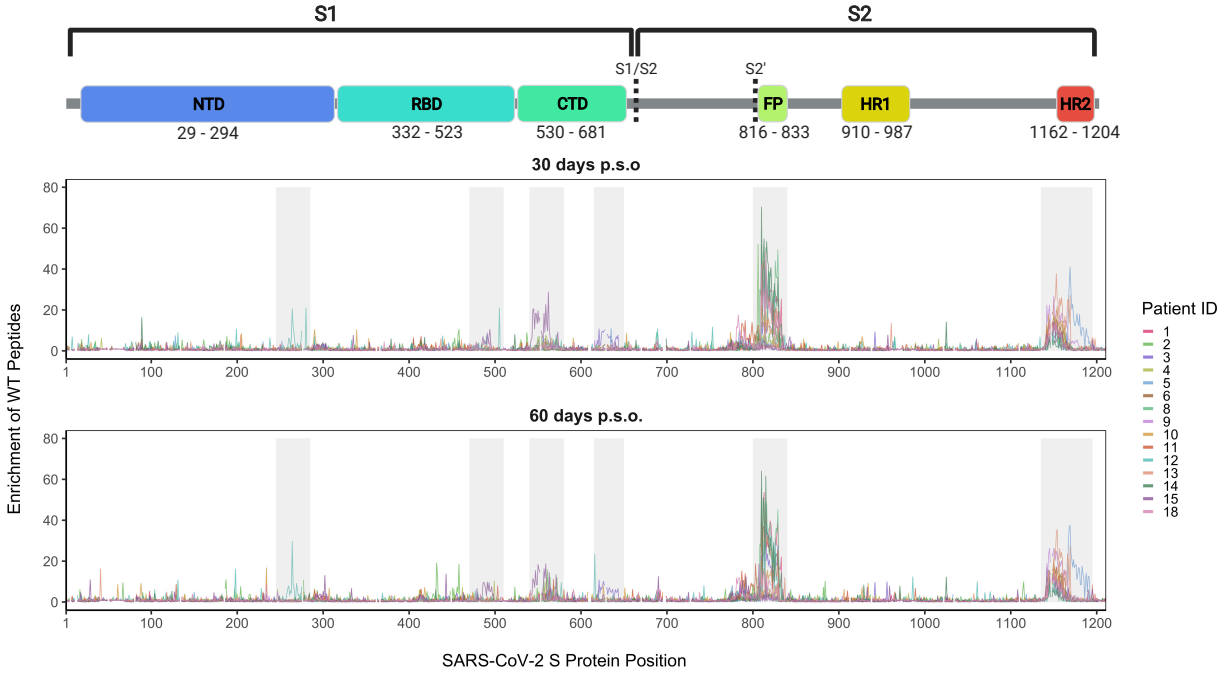


Figure 3.2. Linear epitopes bound by COVID-19 patient plasma. Lines represent the enrichment of wildtype peptides from the Spike Phage-DMS library from individual plasma samples. Samples from convalescent COVID-19 patient plasma taken at approximately day 30 p.s.o. (top panel) or day 60 p.s.o. (bottom panel) are shown. Lines are colored by patient, with the key to the patient IDs on the right (see Supplemental Figure 3.2 for patient inclusion criteria). Grey boxes highlight immunogenic regions where enrichment was detected in at least one individual across timepoints. Peptides that were included in the design, but absent from the phage library (see Supplemental Figure 3.1), are shown as breaks in the line plots. A schematic of S protein domains is shown above, with locations defined based on numbering used in: https://cov.lanl.gov/components/sequence/COV/annt/annt.comp. See also Supplemental Table 3.1 and Supplemental Figures 3.3 and 3.4.

To assess the potential neutralizing capability of these non-RBD targeting antibodies, we depleted the 30 day patient plasma samples of RBD binding antibodies and then performed a neutralization assay with pseudotyped lentivirus expressing the S protein [122]. We performed an ELISA on the mock-depleted and RBD-depleted samples and found very low (12.5% of

mock) to undetectable (0% of mock) residual RBD binding activity (Supplemental Figure 3.3A). We also performed an ELISA against the S protein and found that RBD antibody depletion did not affect binding to the S protein (Supplemental Figure 3.3B), which is consistent with other studies that have found that the majority of S protein-binding antibodies target non-RBD epitopes [61, 123]. We then compared the neutralization capacity of mock-depleted and RBD-depleted patient plasma and found that a wide range (0% to 59%) of the neutralization activity present in patient plasma is directed at non-RBD epitopes (Supplemental Figure 3.3D). While we did not find that enrichment of peptides within the FP or HR2 region correlated with residual neutralization after depletion (Supplemental Figure 3.3E), we did find that binding to the S2 subunit did correlate with NT50 after depletion of RBD antibodies (Supplemental Figures 3.3C and 3.3F). We additionally confirmed that residual antibody binding to RBD after depletion was not associated with the neutralization activity remaining after RBD depletion (Supplemental Figure 3.3G). This suggests that antibodies targeting the S2 subunit are at least partially responsible for the residual neutralization activity in these cases.

Patient-to-patient variability in mutations that lead to loss of antibody binding

To determine the effect of mutations on antibody binding to S protein epitopes, we compared the relative enrichment of wildtype peptides and mutant peptides within the epitopes defined above. To quantify the effect of each amino acid on binding, we calculated the differential selection of mutant peptides versus wildtype peptides and scaled this value by the strength of binding to the wildtype peptide, as we have previously done for Phage-DMS experiments; this measure is highly correlated with the relative binding of individual mutant peptides by ELISA [119]. Plotting the scaled differential selection values for all mutants at each site in a heatmap allows for the visualization of sites where mutations led to a detectable loss (in red) or gain (in blue) of binding. The scaled differential selection data for all patients is also available to view in logo plot form at https://github.com/meghangarrett/Spike-Phage-DMS/tree/master/analysis-and-plotting/dms-view.

We generated heatmaps for four representative samples from COVID-19 patients with strong enrichment. We did this for the two immunodominant regions, FP and linker region/HR2, and included paired day 30 and 60 p.s.o. samples (Figure 3.3 and 3.4). We also examined the escape profiles within less common epitopes (Supplemental Figure 3.4). While the effect of

mutations seemed to be consistent between paired patient samples, between individuals there was more heterogeneity in the specific escape profiles for each region, as follows:

FP region

Within the FP region, sites that led to reduced binding clustered primarily within the 5' region of the FP itself, which spans from sites S816 to F833 (Figure 3.3A). Some sites were similarly sensitive between patients. For example, mutations at site 819 and 820 disrupted antibody binding to some extent in all of the patients shown and were the dominant escape positions for patient 8. In contrast, mutations in the adjacent site 818 appeared to contribute to the epitope of some patient antibodies (1, 3, 12) but not others (8). One interesting site that variably contributed to antibody binding is R815, which is upstream of the FP cleavage site and thus would not be predicted to be present on the post-cleavage form of S2. Antibodies from patient 3 and 12 both showed reduced binding to peptides with mutations at this site, indeed, this was the dominant escape mutation for patient 3. Substitutions at this same amino acid position had a more modest impact on binding and only a few amino acids led to disruption for patient 8. In the case of patient 1, several mutations at site 815 had a positive differential selection value, suggesting they enhance binding.

Not all mutations at each site were equally disruptive to epitope binding, suggesting specific mutations rather than removal of the wildtype amino acid per se may be more important for escape. For example, addition of negatively charged amino acids at site S816 led to loss of binding for patient 8, whereas addition of small amino acids such as alanine or glycine at the same site had little to no effect. We noted some cases where a mutation had a positive differential selection value and thus presumably bound more strongly than the wildtype SARS-CoV-2 sequence at that position. We aligned the FP sequences for SARS-CoV-2 and human endemic coronaviruses and found that some mutations selected above wildtype were residues present in FP sequences from other CoVs (Supplemental Figure 3.5). For patient 12, a few mutants at site 818 including I818L had positive differential selection values, and two coronaviruses (HKU1 and NL63) have a leucine at position 818. In another instance, the N824S mutation had a positive scaled differential selection value for all four samples taken 60 days p.s.o., and we saw that NL63 and 229E both have a serine at position 824.

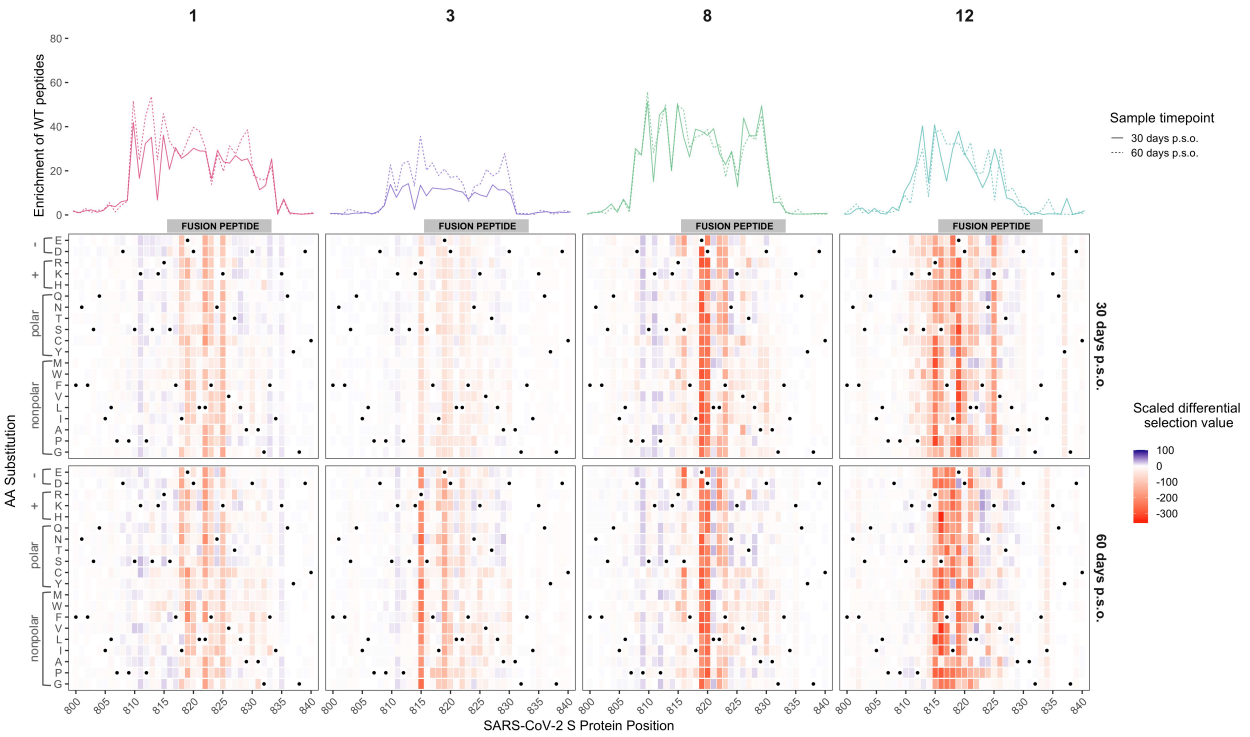


Figure 3.3. Effect of mutations on binding by COVID-19 patient plasma within the FP region. Heatmaps depicting the effect of all mutations, as measured by scaled differential selection, at each site within the FP epitope for representative COVID-19 patients (numbered at top). Mutations enriched above the wildtype residue are colored blue and mutations depleted as compared to the wildtype residue are colored red. The intensity of the colors reflects the amount of differential selection as indicated to the right. The wildtype residue is indicated with a black dot. Line plots showing the enrichment of wildtype peptides for each patient are shown above, with a solid line for the day 30 p.s.o. patient samples and a dashed line for the day 60 p.s.o. patient samples. See Supplemental Figure 3.5 for alignment of coronavirus FP sequences.

Linker/HR2 region

Within the region surrounding HR2, we observed that most patients targeted an epitope spanning aa 1123-1162 within the linker region just upstream of the HR2 domain, with contribution of residues within the N terminus of the HR2 domain in some cases, while one patient targeted an epitope within the HR2 domain itself (Figure 3.4). In addition to this individual variation in the epitope boundaries, the mutations at different sites exhibited marked variability between patients, as seen for the FP epitope. An example of this variability can be seen among patient antibodies targeting distinct epitopes within the linker region. In this epitope, patient 13 exhibited sensitivity to mutations spanning 1162 to 1167, while this was not seen in the other patients. There were also some cases where there were similarities in escape profiles

for patient antibodies targeting the same region. For example, there was evidence that amino acid changes at sites 1151 and 1152 disrupted antibody binding in several patients. However, while mutations at site 1151 disrupted antibody binding for patients 9 and 18, they did not for the antibodies in patient 13.

In patient 5, where there was a distinct epitope within the HR2 domain, we observed unique sensitivity to mutations in sites 1176 through 1182 not found in any other patient. Like many other plasma epitope profiles, mutations to the same site had a variable effect. For example, at site 1180 some mutations improved binding (ex: Q1180G), some reduced binding (ex: Q1180L), while some had little to no effect (ex: Q1180A). There was no overlap in the sites that showed reduced binding in patient 5 and those that were negatively selected in patients 9, 13, or 18, again suggesting that the linker region and HR2 domain epitopes are distinct epitopes that have distinct pathways of escape.

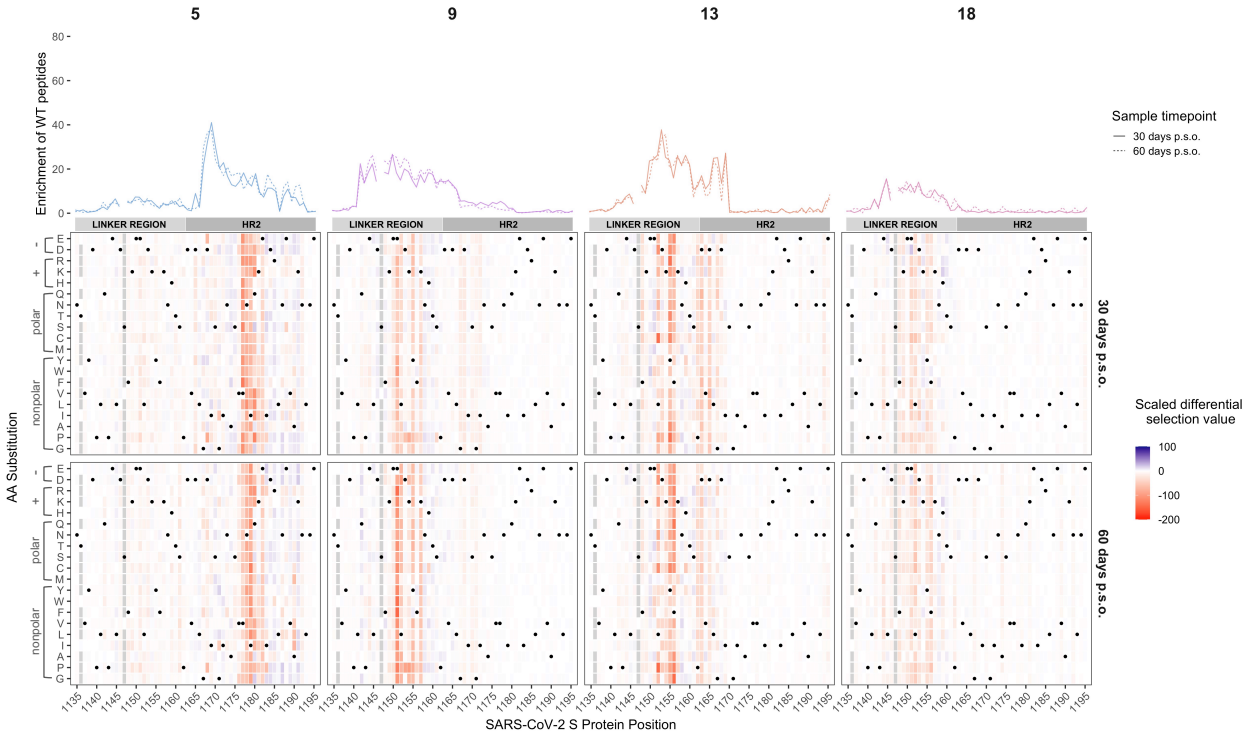


Figure 3.4. Effect of mutations on binding by COVID-19 patient plasma within the linker/HR2 region. Heatmaps depicting the effect of all mutations, as measured by scaled differential selection, at each site within the linker region/HR2 epitope for representative COVID-19 patients (numbered at top). Mutations enriched above the wildtype residue are colored blue and mutations depleted as compared to the wildtype residue are colored red. The intensity of the colors reflects the amount of differential selection as indicated to the right. The wildtype residue is indicated with a black dot. Line plots showing the enrichment of wildtype peptides for each patient are shown above, with a solid line for the day 30 p.s.o. patient samples and a

dashed line for the day 60 p.s.o. patient samples. Peptides missing from the library are shown as grey boxes in the heatmaps and as breaks in the line plots.

Other regions

There was relatively less differential selection at regions outside of the immunodominant epitopes, but the profiles nonetheless provided evidence for some specific mutations that disrupt binding in epitopes in the NTD, RBD, and CTD just downstream of the RBD (Supplemental Figure 3.5). For example, within the NTD, positions spanning 264-269 in patient 12 disrupted binding, most notably mutations from A to select amino acids (W, V and I) at site 264. The escape profiles for patient 15 suggests that changes at positions 491-495 disrupt binding in the RBD epitope, while mutations at sites 550-553 and 558 disrupted binding within the CTD epitope for this patient. For patient 3, the major effect on binding within the CTD was at positions 628-634, most notably at amino acids 628 and 633. Although negative scaled differential selection values can be seen at various other amino acid positions, in many of these cases the effects were weak and/or inconsistent across time points. Identifying cases of stronger enrichment in these regions will be necessary in order to better define the full spectrum of escape mutations at these regions.

Most common circulating SARS-CoV-2 variants, including D614G, are not predicted to escape antibody binding

One question that arises from the escape profiles described above is whether the mutations that disrupt antibody binding identified here correlate with the global evolution of the SARS-CoV-2 virus to-date. Using variant frequencies present in sequences from GISAID and reported at the https://cov.lanl.gov website, we compared the naturally occurring diversity of every site to the effect on antibody binding as found by Phage-DMS. For each patient and at each site, we averaged the scaled differential selection value for all mutants and then plotted this against the mutational entropy, which is a measure of amino acid diversity (Figure 3.5A). Larger mutational entropy values indicate more global diversity at that site, and sites with a mutational entropy value of above 0.02 are flagged as sites of interest by the LANL database.

We found that the majority of sites that led to loss of antibody binding when mutated were not present at a high frequency in nature. Sites within the immunodominant epitopes did not generally have high mutational entropy, although for patient 13 and 15 there were a few sites within the HR2 region that were both present at high frequency in nature (mutational entropy > 0.02) and led to loss of antibody binding when mutated. Averaging the scaled differential selection values across all patients did not reveal any sites that commonly lead to loss of antibody binding and are present at high frequency (Supplemental Figure 3.6).

Focusing on circulating variants that are rising in abundance in the S protein ectodomain, we next examined the effect of mutations at these sites as determined by Phage-DMS. We selected the 20 most common mutations found in nature and saw that these were located across the S1 and S2 proteins and mostly were not found within the major immunodominant regions identified here, with the exception of two mutations that appear within the LR/HR2 region (Figure 3.5B). We examined the scaled differential selection value of each natural variant for all 60 day patient samples. In general, the variants found in the global population did not show reduced binding to patient plasma antibodies, with perhaps a modest effect for a few mutations with individual patient samples. This contrasted with the strong negative selection observed in the Phage-DMS screen for mutations within the two immunodominant epitopes at position 819 and 1156, two positions where there are not variants rising in abundance in nature. However, the possibility remains that sites of high mutational entropy could exist within conformational epitopes, which are not generally displayed in phage libraries.

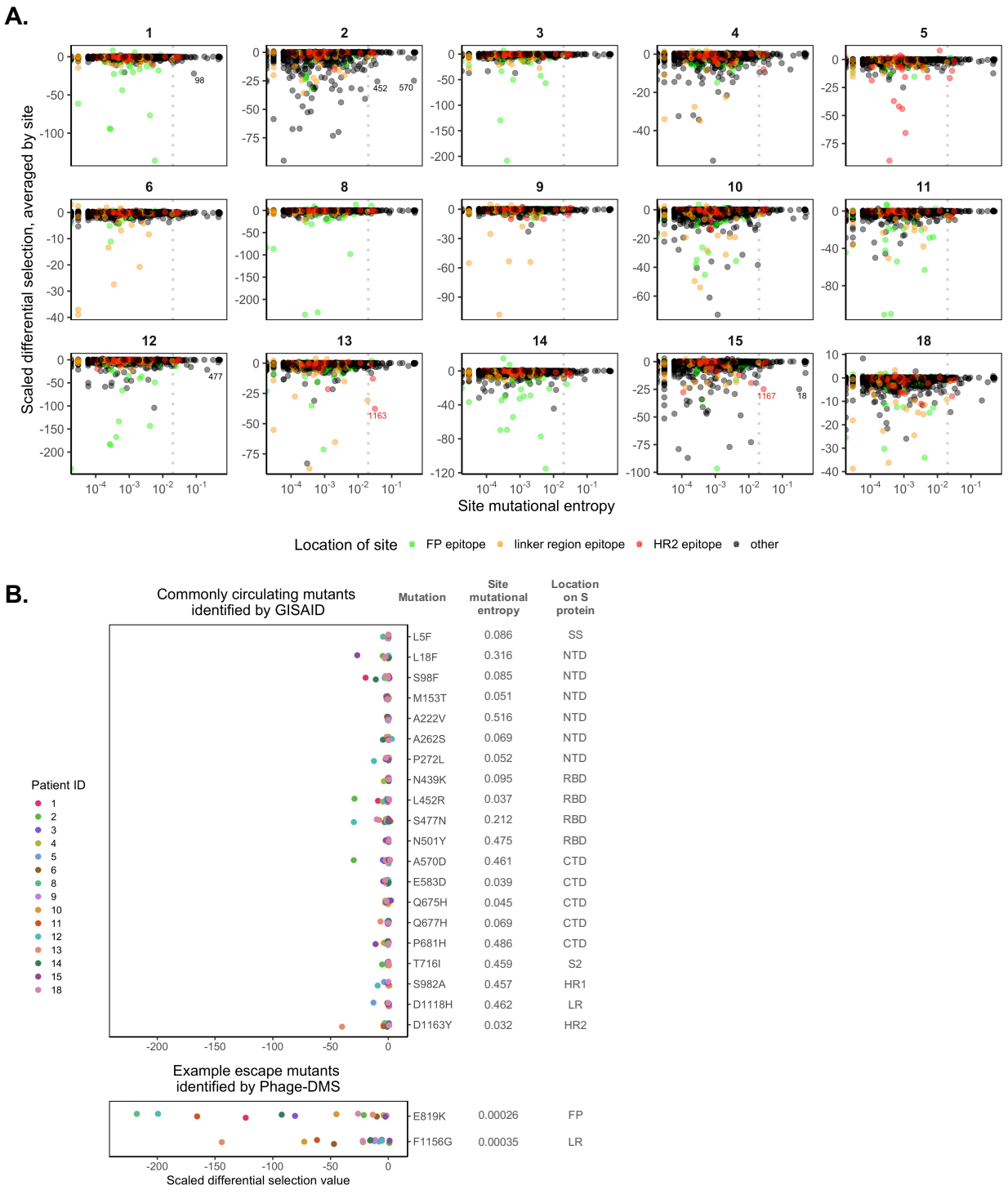


Figure 3.5. Predicted effects of commonly circulating S protein variants on antibody escape. (A) Scatterplot comparing the effect of mutations on patient plasma antibody binding and the frequency of all circulating S protein variants. The mutational entropy of every circulating protein variant, as reported at the https://cov.lanl.gov website and based on GISAID global sequencing, is plotted on the x-axis. The average of the scaled differential selection values for all mutants at each site is plotted on the y-axis. Patient ID's are indicated on the top.

Each site is colored by its location, as indicated on the bottom. The dotted line denotes the cutoff (0.02) of mutational entropy which GISAID uses to determine variants of interest. Sites with relatively high mutational entropy and scaled differential selection values are labelled. See Supplemental Figure 3.6 for the average effect across all patients. (B) Effect of mutant peptides representing commonly circulating S protein variants on binding to COVID-19 patient plasma. We selected sites with a mutational entropy of greater than 0.02, as this is the cutoff used by LANL to determine sites of interest. On top are the 20 sites with the highest mutational entropy values and on the bottom are two selected sites that were noted as sites of antibody escape within immunodominant epitopes by Phage-DMS. On the right are the mutations examined, named according to the wildtype aa, followed by the site number, followed by the mutant aa of interest. Mutations chosen at sites of high mutational entropy represent the most common variant found in nature. The scaled differential values found by Phage-DMS for each mutant peptide are shown as dots and are colored by patient as indicated to the left. Data is from samples taken day 60 p.s.o.. SS = signal sequence, S2 = N-terminal region of S2, LR = linker region.

Evidence for patient-specific epistatic effects of D614G

We were particularly interested in the S protein mutant D614G, a variant that has emerged as the dominant circulating strain of SARS-CoV-2. Because this library also contains peptides made in the background of the D614G strain, we were able to test whether this mutation could, through epistasis, interact with mutations introduced at other sites on the same peptide and together modify the ability of antibodies to bind. For all mutant peptides that tiled across site 614, the enrichment of each peptide made in the original Wuhan Hu-1 strain background (D614) was compared against the enrichment of the corresponding peptide made in the background of the G614 mutation. We found that the presence of the D614G mutation significantly reduced the ability of antibodies from patient 10 to bind to mutant peptides in this region, with a moderate effect size (Figure 3.6, Wilcoxon paired signed-rank test). In 8 out of the other 14 patients there was a statistically significant difference in the binding between mutant peptides with and without the D614G mutation, but the effect sizes were all small. Occasionally, for example in the case of patient 4, a mutation at one position (A609W) showed stronger enrichment in the context of G614 as compared to D614. Conversely, in the case of patient 1, one mutation (A609D) was more enriched when accompanied by D614 as compared to G614. In other cases, such as with patient 2, there was some selective enrichment in both contexts.

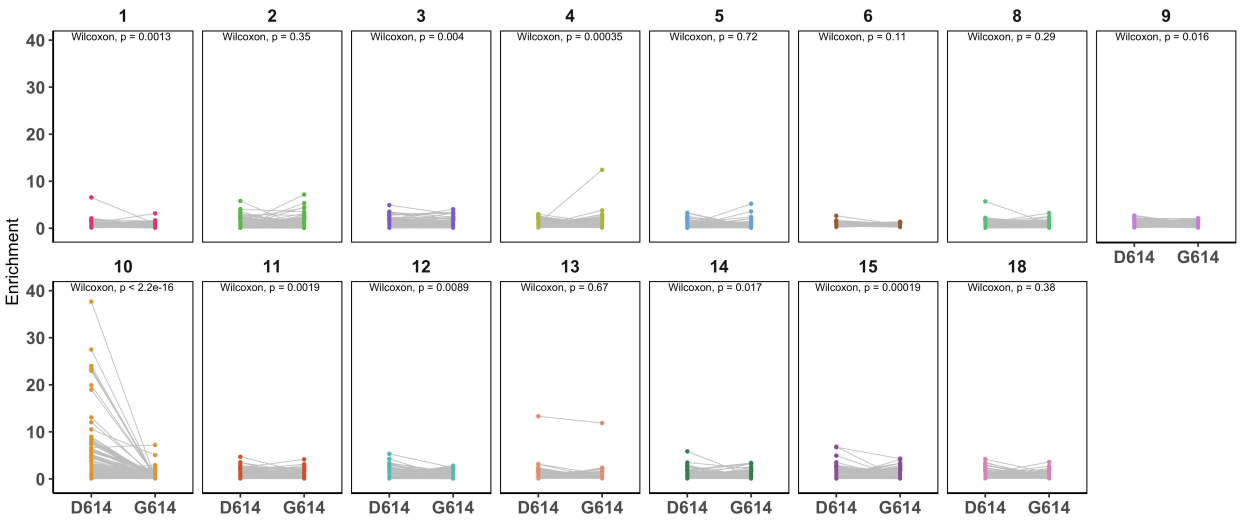


Figure 3.6. Epistatic effects of D614G mutation on antibody binding. Enrichment values for paired mutant peptides made in either the wildtype Wuhan Hu-1 strain (on the left, D614) or D614G background (on the right, G614) for each patient (numbered at top). All mutant peptides that contained site 614 were included in this analysis (spanning aa 599-629). Data is from samples taken day 60 p.s.o.. Wilcoxon paired signed-rank test was performed (n = 380 paired mutant peptides). The effect size for all patient samples was small (Wilcoxon $r < 0.3$) except for patient 10, whose antibodies exhibited a moderate effect (Wilcoxon $r = 0.46$).

3.4 Discussion

In this study, we tested all possible mutations on the S protein to provide a map of escape pathways within immunodominant linear epitopes targeted by the plasma of convalescent COVID-19 patients. The responses within an individual were consistent over time, but there were many unique pathways of escape that differed between individuals, even within the same epitope region. These findings suggest that the pattern of virus evolution within the growing pandemic is not likely to be driven by a single antibody escape mutation, which may explain the lack of emergence of circulating strains with mutations that disrupt antibody binding identified here. Thus, the responses to a SARS-CoV-2 S protein vaccine immunogen are not likely to be uniform, nor will the pathways of escape.

While others have defined the linear epitopes bound by antibodies within COVID-19 patient serum [110, 111, 124-127], there have been limited studies to determine what mutations within these regions could abrogate antibody binding. One study that tested 82 circulating SARS-CoV-2 variants showed that the few single mutations that did result in reduced

neutralization sensitivity were unique to each serum sample [117], and our data supports and extends the idea that antibody binding sites and escape mutants vary greatly from person to person. The lack of a singular escape signature within each immunodominant epitope also implies that a diverse repertoire of antibodies targets these regions. This provides a functional context for results of studies of SARS-CoV-2 S protein antibodies sequences from COVID-19 patients, which showed that no single clones dominate the antibody response; rather, a diverse collection of variable heavy and light chain genes are used in different individuals [128, 129].

Our results also demonstrate the power of interrogating the role of every possible amino acid at every site on the S protein. Mapping COVID-19 patient serum epitopes by alanine scanning has helped identify sites of antibody binding by removing important side chains interactions [125], but studies with other viruses have shown that escape can be mediated by mutations at sites not directly in contact with the antibody via introduction of nearby charged or bulky amino acids [57, 59, 130]. This concept was evident in this study, for example within the FP epitope for patient 8, where addition of negatively charged amino acids led to escape at site S816 but addition of an alanine did not. We were also able to examine the effects of mutations in the context of the common circulating 614 variant, which we found did not itself drive escape from antibody binding. However, our data suggests that mutation D614G does potentiate escape mutations at other positions in at least one patient, highlighting the power of examining combinations of mutations to better understand the increasing global dominance of the D614G variant.

The FP and linker region/HR2 immunodominant epitopes profiled in this study offer intriguing alternative targets for vaccine design or immunotherapy that could complement efforts focused on the RBD. S2 in general, and the FP in particular, is highly conserved among coronaviruses, indicating the strong purifying evolution acting in this region due to functional constraints. As with other RNA viruses, this restricted mutational space likely limits the virus' ability to escape antibody binding [131-134]. One study has found that the FP is a target of neutralizing antibodies in SARS-CoV-2 patients [110], and these antibodies could act by preventing protease-mediated cleavage at the S2' site. Interestingly, we saw evidence of antibody binding to sites upstream of the S2' site in some patients (3 and 12), but not others (1 and 8), potentially indicating that antibodies to the FP could bind during different stages of cell entry, before or after S2' cleavage. HR2 and the upstream linker region are both targets of neutralizing antibodies in SARS-CoV-2 patients [111], and a neutralizing antibody raised against the linker region from murine SARS-CoV infection has been shown to cross-react with SARS-

CoV-2 [135, 136]. Antibodies targeting the heptad repeat regions, which undergo large conformational transformations in order to facilitate membrane fusion, could neutralize by binding to the fusion intermediate state. Interestingly, our data suggest that non-RBD directed neutralization correlates with S2 protein binding, although it did not correlate specifically with either the FP or linker region/HR2 enrichment measured by Phage-DMS. These more conserved targets in S2 may be important for designing optimal and durable vaccines, given that RBD has higher mutational entropy, increasing the potential for immune escape from vaccine-induced antibodies. The mutational flexibility of the RBD is supported by *in vitro* studies with RBD-targeting neutralizing antibodies, which found that escape mutants in the RBD of SARS-CoV-2 were rapidly selected [137].

There are several important caveats to the results we obtained with Phage-DMS libraries. In general, phage libraries display linear peptides and therefore miss antibodies that bind to complex conformational epitopes, or at best provide a partial view of a linear portion of such epitopes. The strong plasma binding response to RBD by ELISA yet lack of enriched RBD peptides demonstrates this limitation and indicates that we are unable to make predictions about mutants in conformational epitopes on the RBD. Additionally, the peptide libraries are amplified in bacteria and therefore lack glycans or other post-translational modifications. While these Phage-DMS-derived escape maps define mutations that could lead to loss of antibody binding, it is unknown whether the virus would tolerate mutations at those sites. In fact, a study examining the effect of mutations within the SARS-CoV-2 RBD demonstrated that many led to poor protein expression or loss of function [118]. Thus, for the antibody targeting regions described here, which are in more conserved functional domains, the escape mutations should be further examined in the context of their mutational tolerance.

These studies have defined common and variable escape mutations across 18 COVID-19 patients that will be useful for viral surveillance, particularly as SARS-CoV-2 S protein-based vaccines are introduced into the population. In addition, the Spike Phage-DMS library developed here could be useful for examining larger cohorts, potentially including those with variable clinical outcomes and individuals of variable ages, to define whether mutations that disrupt antibody binding vary in a systematic way across populations and whether this is correlated with clinical outcome or risk of reinfection.

Limitations of the Study

This study focuses on defining the epitopes, including linear sequences, targeted by convalescent plasma in mild and moderate cases of COVID-19. Specifically, because the Spike Phage-DMS library displays only short linear peptides, this method is generally not optimal for identifying the full epitope of antibodies that rely on conformational presentation of the antigen. The peptides displayed in a Phage-DMS library do not undergo post-translational modification and therefore antibodies that rely on binding to glycans and other similar moieties cannot be defined using this method. The plasma samples used in this study were taken relatively soon after infection and were mostly taken from patients with mild infection not requiring hospitalization. Thus we may not have detected the full range of possible responses to COVID-19.

3.5 Methods

RESOURCE AVAILABILITY

Materials Availability

Further information and request for reagents may be directed to the corresponding author Julie Overbaugh (joverbau@fredhutch.org).

Data and Code Availability

Sequencing data has been deposited to NCBI and are accessible under BioProject # PRJNA715823. The Nextflow pipeline, used to align all sample reads to the reference library, is available at https://github.com/matsengrp/phip-flow. A custom python package used for all downstream sample curation and analysis is available at https://github.com/matsengrp/phippery. All custom code and input files used to (1) generate oligonucleotide sequences for the S protein Phage-DMS library, (2) run the alignment pipeline, (3) analyze sequencing data for the experiments in this paper, and (4) to generate figures have been deposited at https://github.com/meghangarrett/Spike-Phage-DMS. Data is available to explore using the dms-view website, hosted here: https://github.com/meghangarrett/Spike-Phage-DMS/tree/master/analysis-and-plotting/dms-view.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Subjects

Plasma samples were taken from non-hospitalized COVID-19 patients enrolled in the Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) study at the University of Washington. Prior to study initiation, the following institutional human subjects review committee approved the protocol: University of Washington IRB (Seattle, Washington, USA) and concurrent approvals were obtained from the Fred Hutchinson Cancer Research Center for the current study. All plasma samples were heat-inactivated at 56 ºC for 1 hour before storage and use. Plasma samples were spun in a centrifuge for 10 min at 1,000 x g in order to clarify the supernatant before use.

METHOD DETAILS

Design and generation of the Spike Phage-DMS library

To create a Phage-DMS library for the S protein of SARS-CoV-2 we used the sequence from the Wuhan Hu-1 strain (Genbank: MN908947). Only the ectodomain of the S protein was included (aa 1-1211), excluding the transmembrane and cytoplasmic domains. Additionally, in the region of the D614G variant of the Wuhan Hu-1 strain, we designed DMS peptides spanning positions 599 through 619 that also include the D614G mutation. Sequences were optimized for uniform GC content (to reduce later biases during PCR amplification) and codon usage for expression in *E coli*. GC and codon optimization was done using the Python package DNAChisel (version 3.2.2), aiming for GC content of between 0.4 and 0.6 within a window of 100 nucleotides. After optimizing the sequences, the two subunits of the S protein (S1 and S2) were then treated as separate proteins. Sequence coding for a glycine-serine linker ($[G_4S]_3$) was added to the beginning and end of the sequence of each protein in order to ensure that the first amino acid of the protein was located in the central position of the peptide. We then generated sequences coding for peptides 31 amino acids long, tiling by 30 amino acids, and containing a single variable residue at the central position of the peptide. This resulted in 20 peptide

sequences containing all possible mutations at each position along the proteins, with only one amino acid shift between sequential peptides. Each sequence additionally had 5' and 3' adaptor sequences added to facilitate amplification and cloning (5': AGGAATTCTACGCTGAGT, 3': TGATAGCAAGCTTGCC). After removing duplicate sequences, 24,820 unique sequences were synthesized by Twist Bioscience as an oligonucleotide pool. Two biological duplicate libraries were generated by independently cloning the sequences into a T7 phage vector and then amplifying the phage, as we have done previously [119]. Peptides are numbered by the corresponding S protein location of the amino acid in the central position of the peptide. Code used to optimize and generate peptide sequences for the S protein Phage-DMS library can be found at https://github.com/meghangarrett/Spike-Phage-DMS.

Immunoprecipitation of human plasma with Phage-DMS library

Immunoprecipitation of phage-antibody complexes was performed as previously described [63, 119]. Briefly, deep 96-well plates were blocked with 3% BSA in Tris-buffered saline with 0.01% Tween (TBST) overnight at 4°C. The phage library was diluted to a concentration representing 200,000 pfu/mL per unique peptide and 1 mL of the diluted phage was added to each well. We assume that plasma IgG concentrations are about 10 ug/uL [138] and added 10 ug of each sample to the appropriate wells. For every experiment, samples are run in technical duplicate on the same plate. Plates are sealed and rocked at 4°C for 18-20 hours. To immunoprecipitate the phage-antibody complexes we added 40uL of a 1:1 mixture of Protein A and Protein G Dynabeads to each well and incubated the samples for 4 hours at 4°C while rocking. Dynabeads were magnetically separated and then beads were washed 3x with 400 uL wash buffer (150 mM NaCl, 50 mM Tris-HCl, 0.1% [vol/vol] NP-40, pH 7.5). We resuspended beads in 40 uL of water and then lysed bound phage at 95°C for 10 minutes. Additionally, we lysed 10-20 million phage from the diluted input library to determine the distribution of phage in the starting library. Lysed samples were stored at -20°C before preparing for Illumina sequencing.

Illumina library preparation and deep sequencing

Lysed phage DNA from each sample was amplified and readied for Illumina deep sequencing by performing two rounds of PCR, as previously described [119]. Each PCR reaction was performed using Q5 High-Fidelity 2X Master Mix. For the first round of PCR, 10uL of lysed

phage was used as the template in a 25 uL reaction. For the second round of PCR, 2 uL of the round 1 PCR product was then used as the template in a 50 uL reaction, with primers that add dual indexing sequences on either side of the insert. PCR products were then cleaned using AMPure XP beads and eluted in 50 uL water. DNA concentrations were quantified via Quant-iT PicoGreen dsDNA Assay Kit. Equimolar amounts of DNA from the samples, along with 10X the amount of the input library samples, was pooled, gel purified, and the final library was quantified using the KAPA Library Quantification Kit. Pools were sequenced on an Illumina MiSeq with 1x125 bp single end reads using a custom sequencing primer.

Depletion of RBD-binding antibodies from patient plasma

Depletion was performed as previously described [61]. In brief, magnetic beads conjugated with RBD protein were washed in PBS with 0.05% BSA and resuspended at 1mg/mL. Beads and plasma were incubated together at a ratio of 1:3 plasma:beads, and plasma was allowed to bind to beads overnight on a rotating rack at 4°C. Beads were then magnetically separated and the plasma supernatant was used in downstream assays (accounting for the 1:4 dilution factor that occurred during depletion). A mock depletion control for each plasma sample was also prepared in parallel, adding PBS with 0.05% BSA at a ratio of 1:3 plasma:buffer.

ELISA with RBD, Spike, and S2 proteins

ELISAs to test for plasma IgG binding to RBD, Spike, and S2 proteins were performed as previously described [139, 140], with some modifications. 384 well ELISA plates were coated with 25 uL of recombinant protein at a concentration of 2 ug/mL in phosphate buffered saline (PBS), with protein allowed to coat the wells overnight at 4°C. Plates were then washed 4x with 100 uL of wash buffer (PBS with 0.01% Tween-20 [PBST]) and blocked with 100 uL of 3% nonfat dry milk for 2 hours at room temperature. Block was then thrown off, and 25 uL of diluted plasma or antibody control was added to each well. Patient plasma was diluted 1:100 (accounting for prior dilution made during RBD-antibody depletion) in wash buffer with 1% nonfat dry milk and then four 3-fold serial dilutions were made, for a total of 5 dilutions. CR3022 was used as a positive control for RBD and Spike and was diluted to 1 ug/mL followed by nine 4-fold serial dilutions, for a total of 10 dilutions. Pooled plasma from the 18 patients in this study was used as the positive control for S2 and was diluted 1:50 followed by nine 4-fold serial

dilutions, for a total of 10 dilutions. Pooled plasma collected pre-pandemic was used as a negative control, and was treated the same as patient plasma. Plates were incubated for 2 hours at 37°C and then samples were washed 4x with wash buffer. Goat anti-human IgG-Fc horseradish peroxidase (HRP)-conjugated antibody was diluted 1:3000 in wash buffer containing 1% nonfat dry milk, and 25 uL was added to the plate. After 1 hour at room temperature, plates were washed 4x with wash buffer and 25 uL of TMB substrate was added. After 15 minutes at room temperature, the reaction was stopped with 25 uL of 1N sulfuric acid and the OD450 was read on a BioTek Epoch plate reader. After subtracting background signal from buffer-only wells, the area under the titration curve was calculated using log-transformed dilution values in GraphPad Prism (v9).

Generation and titration of Spike pseudotyped lentivirus

We generated and determined the titers of pseudotyped lentivirus as previously described [122]. We used the codon-optimized Spike sequence from the Wuhan-Hu-1 strain with a 21 amino acid deletion in the cytoplasmic tail (also known as HDM_Spikedelta21). HEK293T cells were first seeded at a density of 5 x 10^5 cells per well into 6-well plates with DMEM (supplemented with 10% fetal bovine serum (FBS), 2 mM l-glutamine, and penicillin/streptomycin/fungizone). After 16 to 24 hours, cells were transfected using FuGENE-6 with the Luciferase_IRES_ZsGreen backbone, the Gag/Pol-, Rev-, and Tat-containing lentiviral helper plasmids, as well as the plasmid containing Spike. 24 hours after transfection, we removed the media and replaced with fresh DMEM complete. At 50-60 hours post transfection we collected viral supernatants, filtered through a 0.22 um Steriflip filter, and stored at -80°C. To titer the viral supernatant, we plated 1.25 x 10^4 HEK293T-ACE2 cells per well in 50 uL in a 96-well black-walled plate and incubated 16 to 24 hours before adding 100 uL of viral supernatant dilutions to each well. Viral supernatants were diluted 1:10 in DMEM complete followed by seven 2-fold serial dilutions, and each dilution series was run in duplicate. 60 hours post-infection we removed 100 uL of media from each well and added 30 uL of the Bright-Glo reagent, and then read the relative luciferase units (RLU) on the LUMIstar Omega plate reader.

Neutralization assays

We performed neutralization assays as previously described [122]. We plated 1.25 x 10^4 HEK293T-ACE2 cells per well in 50 uL in a 96-well black-walled plate and incubated for 12 hours before infecting. Virus was diluted to achieve a final concentration of ~200,000 to 500,000 RLU/well, which was determined by titering as described above. Starting at 1:20 (accounting for final virus/plasma volume), six 3-fold serial dilutions of patient plasma, for a total of 7 dilutions, were made and then added to virus. The virus/plasma mixture was incubated at 37°C for 1 hour and then 100 uL was added to the cells. At 60 hours post infection we measured the RLU per well using the Bright-Glo reagent on the LUMIstar Omega plate reader. We calculated the fraction infectivity of the antibody-containing wells by normalizing to a media-only (no plasma) well infected with the same diluted virus. We calculated the neutralization titer 50% (NT50) using GraphPad Prism software by fitting a four-parameter nonlinear regression curve with the bottom fixed at 0 and the top fixed at 1.

Multiple sequence alignment

Alignment of FP amino acid sequences from SARS-CoV-2, OC43, HKU1, NL63, and 229E was done using Clustal Omega. GenBank sequences used are as follows: YP_009724390.1, YP_009555241.1, YP_173238.1, YP_003767.1, and NP_073551.1.

Graphical illustrations

Graphical abstract, Figure 3.1, and the diagram of S protein domains in Figure 3.2 were made in BioRender.com

QUANTIFICATION AND STATISTICAL ANALYSIS

Demultiplexing and alignment of Illumina reads

Demulitplexing and fastq file generation were performed by the Fred Hutch Genomics Core using Illumina MiSeq Reporter software. Demultiplexed sample reads were aligned to the reference library in parallel using a Nextflow data processing pipeline. The pipeline builds a Bowtie index from the peptide metadata by converting the metadata to fasta format and feeds it into the bowtie-build command. The low-quality end of the reads is trimmed to 93bp in order to match the reference lengths before performing end-to-end alignment and allowing for 0

mismatches. For each sample, we quantified the abundance of each peptide by using samtools-idxstats to count the number of reads mapped to each specific peptide in the reference library. The peptide counts were merged into an enrichment matrix organized by unique identifiers for each peptide and sample. The metadata tables were tied with the enrichment matrix into an xarray dataset using shared coordinate dimensions of the unique sample and peptide identifiers. We used this dataset organization as the starting point for all downstream sample curation and analysis.

Calculating enrichment and scaled differential selection of peptides

Enrichment and scaled differential selection were calculated as described previously [119]. To calculate enrichment, each peptide's pseudocount frequency was divided by a respective library pseudocount. Each pseudo count was defined as the raw count plus the ratio of the sum of each sample and library count, with a minimum value of P=1. Differential selection of mutant amino acids was then calculated as the log-fold change between each mutant peptide and the wildtype peptide at a locus. By definition, the differential selection of a wildtype amino acid is always 0. The scaled differential selection of a mutant amino acid was then calculated by multiplying the differential selection of a mutant amino acid by the enrichment of the wildtype peptide at that position.

Data curation

All samples were screened at least once with each Spike Phage-DMS Library 1 and Library 2, but occasionally were screened multiple times with each library. To ensure we used the same amount of data with each sample, we examined the Pearson's correlation value of enrichment values for all peptides between all biological replicate experiments and chose the data from the best two correlated experiments for inclusion in the analyses presented here. If there was not at least one timepoint where the sample had a correlation of at least 0.5, then all samples from that patient were excluded from enrichment and scaled differential selection analyses. Peptides that were never sequenced in any experiment were removed from the analyses, and all enrichment and scaled differential selection values shown are the average of two biological replicate experiments.

3.6 Notes

ACKNOWLEDGMENTS

We gratefully acknowledge Kevin Sung for helpful discussion and Katherine H. D. Crawford for Python code used to generate the oligonucleotide library sequences. We thank Sarah K. Hilton, John Huddleston, and Jesse D. Bloom for developing and demonstrating the dms-view online tool. We thank the Fred Hutch Genomics core facility, and in particular Cassie Sather, for assistance with sequencing. This work was funded by NIH grants AI138709 (PI Overbaugh), R01 AI146028 and U19 AI128914 (PI Matsen). Julie Overbaugh received support as the Endowed Chair for Graduate Education (FHCRC). The research of Frederick Matsen was supported in part by a Faculty Scholar grant from the Howard Hughes Medical Institute and the Simons Foundation.

AUTHOR CONTRIBUTIONS

JO conceived the project; MEG and JO led the design of the study; HYC led the HAARVI study, with CRW, JKL, and DM involved in sample collection. MEG, HLI, and HW performed experiments, with CIS performing sample processing and organization; MEG and JG performed computational and data analyses, with FAM advising. MEG and JO wrote the paper with input from all authors.

DECLARATION OF INTERESTS

MEG and JO are inventors on a patent application on Phage-DMS. HYC is a consultant for Merck, Pfizer, Ellume, and Bill and Melinda Gates Foundation and has received support from Cepheid and Sanofi-Pasteur.

INCLUSION AND DIVERSITY STATEMENT

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. We worked to ensure that the study questionnaires were prepared in an inclusive way. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community. One or more of the authors of this paper self-identifies as living with a disability. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work

Chapter 4

SPIKE EPITOPES DIFFER BETWEEN ANTIBODIES ELICITED BY SARS-COV-2 MRNA VACCINES VERSUS INFECTION

Sections of text in this chapter have been modified slightly from a manuscript currently in submission:

Garrett, M.E., Galloway, J., Wolf, C.R., Logue, J.K., Franko, N., Weight, H., Chu, H.Y., Matsen IV, F.A., Overbaugh, J. Spike epitopes differ between antibodies elicited by SARS-CoV-2 mRNA vaccines versus infection. (2021). *In submission*.

4.1 Abstract

Control of the COVID-19 pandemic will rely on SARS-CoV-2 vaccine-elicited antibodies to protect against emerging and future variants; an understanding of the differential humoral responses from infection or vaccination is needed to achieve this goal. Here we comprehensively profiled the epitopes and pathways of escape for Spike-specific antibodies in individuals with diverse infection and vaccination history. Individuals with mild infection had antibodies that bound to epitopes in the S2 subunit (fusion peptide and heptad-repeat regions), whereas those with severe infection or vaccination had antibodies that additionally bound to epitopes in the S1 subunit (N- and C-terminal domains). Epitope binding appeared to change over time after vaccination, but other covariates such as mRNA vaccine dose, mRNA vaccine type, and age did not affect antibody binding to these epitopes. Vaccination induced a uniform escape profile across individuals in many cases. These findings shed light on potential SARS-CoV-2 escape mutations selected for at the population level.

4.2 Introduction

The future of the COVID-19 pandemic will be determined in large part by the ability of vaccine-elicited immunity to protect against current and future variants of the SARS-CoV-2 virus. Several vaccines have now been approved for use in multiple countries, including two that are based on mRNA technology: BNT162b2 (Pfizer/BioNTech) and mRNA-1273 (Moderna). In the United States, over half of adults are now vaccinated against SARS-CoV-2, the majority of whom have received one of these mRNA vaccines. While these vaccines have been shown to

effectively guard against infection, severe disease, and death related to SARS-CoV-2[141-147], less is known about how effective they will be against emerging and future variants. Current surges in the Delta variant coupled with reports of reduced potency of vaccine elicited antibodies against this variant highlight this concerning ongoing dynamic[148, 149]. Evidence from related endemic coronaviruses indicates that evolution in the Spike protein results in escape from neutralizing antibodies elicited by prior infection[52], potentially contributing to why endemic coronaviruses can reinfect the same host[53, 150, 151]. Without immunity that is robust in the face of antigenic drift, continual updates of the vaccine to combat new SARS-CoV-2 variants will likely be necessary to provide optimal protection against symptomatic infection.

Prior infection with SARS-CoV-2 also provides some immunity against subsequent re-infection, and several studies have characterized the epitopes targeted by convalescent sera[47, 61, 125, 152, 153]. It is currently unknown whether SARS-CoV-2 infection and vaccination result in antibodies that bind to similar epitopes, an important point to consider given that most people have acquired antibodies through immunization and not infection. The Spike protein encoded by the mRNA in both SARS-CoV-2 vaccines is stabilized in the prefusion conformation by addition of two proline substitutions[20]. This change in sequence and fixed conformation of the Spike protein could result in altered antibody targeting when compared to antibodies elicited during infection, where Spike undergoes several conformational changes. It is also possible that differences in antibody specificity could be due to the amount of antigen or type of immune response stimulated in the context of infection versus vaccination. We know that vaccines drive higher neutralization titers and more Spike binding IgG antibodies than infection[154-156], indicating some differences in the B cell response compared to infection. A recent study showed that antibodies against the receptor binding domain (RBD) of Spike differ between infected and vaccinated individuals; they are generally less sensitive to mutation and bind more broadly across the domain in the context of vaccination as compared to infection[62].

Although the majority of the serum binding response in SARS-CoV-2 infected and vaccinated people is directed towards regions of the protein outside of the RBD epitopes[61, 62, 123, 157], few studies have examined the prevalence and escape pathways of these antibodies, especially in the setting of vaccination. Antibodies to linear epitopes in the S2 domain of Spike overlapping the fusion peptide (FP), and in the linker region just upstream of heptad repeat 2 (L-HR2) region are found in serum from COVID-19 patients, and some studies suggest these antibodies may be neutralizing[110, 111]. These non-RBD responses may also be important contributors to non-neutralizing antibody activities, which have been associated with protection

and therapeutic benefit in experimental SARS-CoV-2 models and with vaccine protection[44, 45, 51, 158]. Importantly, these epitopes lie in more conserved regions of Spike than RBD where functional constraints on variation may counter the selective pressure for viral escape.

To compare antibody immunity elicited by SARS-CoV-2 infection and vaccination, we used a high-resolution Spike-specific deep mutational scanning phage display library to profile the epitopes and sites of escape for serum antibodies from people who had been infected, vaccinated, or a combination of both. This approach, called Phage-DMS, identified four non-RBD antibody binding epitopes across all samples: the FP and L-HR2 region in the S2 subunit, and the N-terminal and C-terminal domains (NTD and CTD, respectively) in the S1 subunit of Spike. Antibodies to NTD and CTD were uniquely present in the setting of mRNA vaccination or severe infection, but mostly absent in mild COVID-19 cases. In vaccinated individuals, the magnitude of the response varied over time both to the CTD and L-HR2 epitopes. Other covariates, such as age, dose, and vaccine type had no significant differences in the binding profiles observed. Of particular relevance to protection against emerging variants, infection and vaccination appear to shape the pathways of escape differently in different epitopes. In the FP epitope, which is a dominant response after infection, the escape pathway was maintained after subsequent vaccination; in the L-HR2 epitope, infection resulted in antibodies with diverse pathways of escape, whereas vaccination induced a highly uniform escape profile across individuals. Overall, these findings indicate that vaccination induced a broader antibody response across the Spike protein but induced a singular antibody response at the L-HR2 epitope, which could favor variants that emerge with these mutations.

4.3 Results

Samples from individuals with varying SARS-CoV-2 infection and mRNA vaccination histories profiled using high resolution Spike Phage-DMS library

We collected serum samples from two cohorts, termed the Moderna Trial Cohort and the Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) Cohort[157, 159]. The Moderna Trial Cohort were participants in a Phase 1 trial and consisted of 49 individuals, 34 who received the 100 μg dose of mRNA-1273 (Moderna) and 15 who received the 250 μg dose. Serum samples were taken at days 36 and 119 post first dose (7 and 90 days post second dose, respectively[159]. Serum samples were taken at days 36 and 119 post first dose

(7 and 90 days post second dose, respectively)[159]. The HAARVI Cohort included 64 individuals, 44 who had confirmed SARS-CoV-2 infection and 20 who had no reported infection; among this group, 44 were also vaccinated. Those with infection history were stratified by severity based on hospitalization status (39 non-hospitalized/mild vs. 5 hospitalized/severe) and serum was sampled at timepoints ranging from 8 to 309 days post symptom onset. Of these 44 individuals, 24 were also sampled after vaccination with two doses of either mRNA-1273 (Moderna, n=8) or BNT162b2 (Pfizer/BioNTech, n=15), with 23 from the non-hospitalized group and 1 from the hospitalized group. All 20 SARS-CoV-2 naïve individuals were sampled post-vaccination, with 18 having an additional sample taken pre-vaccination (0 to 98 days). Post-vaccination timepoints for all naïve and convalescent individuals ranged from 23 to 65 days after the first dose (5 to 42 days after the second dose, respectively). Figure 4.1 provides an illustration of the two cohorts and their respective samples' infection and vaccination statuses.

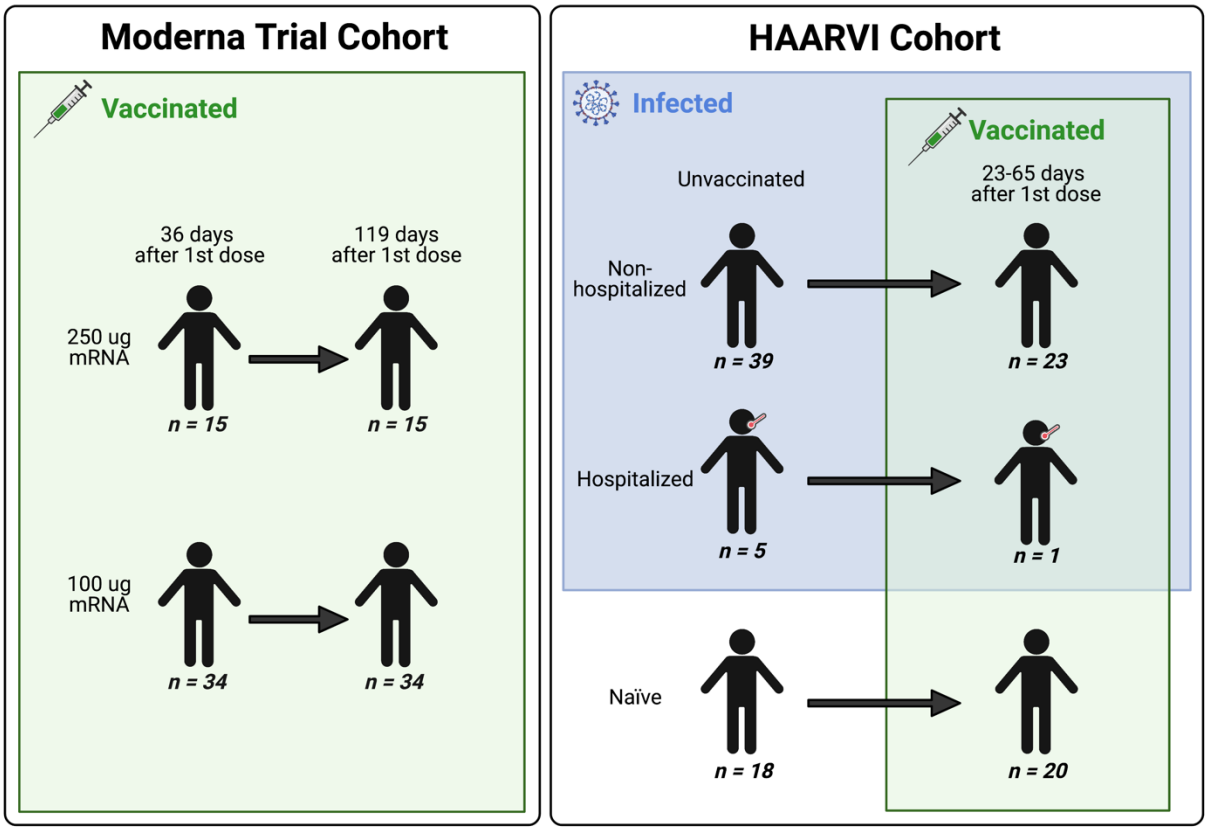


FIGURE 4.1: A schematic of sample cohorts. Characteristics of individual participants sampled as part of the Moderna Trial Cohort (left) or the HAARVI Cohort (right). Sample sizes of unique individuals in each group are designated below each figure.

We used a previously described Spike Phage-DMS library to profile the epitopes bound by serum antibodies in the samples described above[157]. This library consists of peptides displayed on the surface of T7 bacteriophage that are 31 amino acids long, tiling across the length of Spike in one amino acid increments. Peptides in the library correspond to the wild-type Wuhan Hu-1 Spike sequence as well as sequences that contain every possible single amino acid mutation at the central position of the peptide. Serum samples were screened with this library by performing immunoprecipitation (IP) followed by sequencing of the pool of phage enriched by the serum antibodies as previously described[63, 119, 157].

Serum antibodies bind to distinct epitopes in infected and vaccinated individuals

We first examined the wild-type peptides in the Spike Phage-DMS library that were enriched by each serum sample to determine the epitopes bound by antibodies in each sample from these cohorts (Figure 4.2A). The major targeted epitopes across all the cohorts were in the NTD, CTD, FP, and L-HR2 regions. Serum from non-vaccinated infected individuals who were not hospitalized mostly bound to immunodominant epitopes in the FP and L-HR2, both of which are epitopes previously identified in infected individuals using Phage-DMS[157]. Samples from hospitalized/severe COVID-19 cases and vaccinated individuals also bound to the FP and L-HR2 regions, but additionally bound to epitopes within the NTD and CTD regions. In naïve serum samples there were antibodies that occasionally bound to the FP and L-HR2 peptides. These findings likely reflect that some individuals have preexisting cross-reactive antibodies that bind to these conserved regions between SARS-CoV-2 and endemic coronaviruses, as suggested by previous studies[125, 153].

A Principal Component Analysis (PCA) was used to further investigate differences between the infected and/or vaccinated groups. This analysis indicated that binding to epitopes in the NTD, CTD, FP, and L-HR2 regions were driving differences between samples (Figure 4.2B). To quantify differences in antibody binding between groups, for each sample we summed together the enrichment values within each identified epitope region and performed pairwise comparisons between non-hospitalized infected people and all other groups (Figure 4.2C). Most strikingly, we found non-trivial group differences in the magnitude of humoral responses to these major epitopes on the Spike protein. Specifically, antibodies from both hospitalized infected and vaccinated individuals had significantly higher binding to the NTD, CTD, and L-HR2 regions compared to non-hospitalized infected individuals. However, antibodies from non-hospitalized

infected individuals displayed significantly higher binding to the FP epitope than samples from hospitalized or vaccinated individuals. There was no significant difference in any epitope binding in these four regions between vaccinated samples with and without prior infection (p>0.05, Mann-Whitney-Wilcoxon [M.W.W.]).

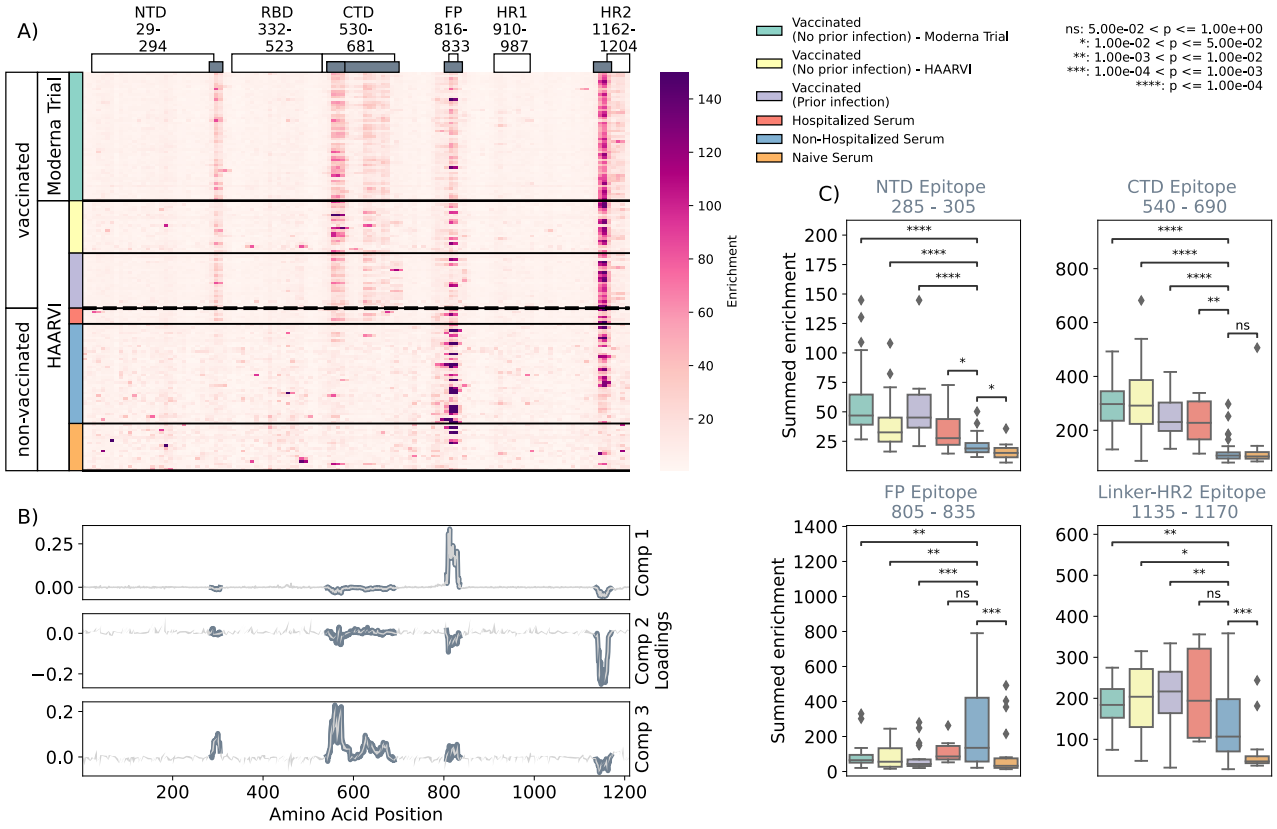


FIGURE 4.2: Enrichment of wild-type peptides by serum antibodies. (A) Heatmap with a sample in each row and groups of samples colored on the left. Columns represent peptide locations, with each square on the heatmap indicating the summed enrichment value within a 10-peptide interval. Darker purple indicates higher enrichment values, and values above 150 were capped. Transparent boxes above the heatmap annotate the Spike protein domains, while the smaller grey boxes indicate *epitope binding regions* defined in this analysis (B) The loading vectors from the PCA analysis with the four epitope sites highlighted; enrichments in each of these regions are summed together for subsequent analysis. (C) Box plots describing the distribution of summed wild-type enrichment values for each sample within each of the four epitope sites, each named according to its associated protein domain. Color indicates the sample group. The bars between boxplots give statistical significance (p-value) tests using a Mann-Whitney-Wilcoxon test. All sample group comparisons with the non-hospitalized infected group were performed, and only significant values are shown.

Effect of age, dose, vaccine type, and timepoint on epitope binding

In order to determine if there were covariates that contributed to differences in antibody binding, we examined the effect of participant age, vaccine dose and type, and timepoint post infection or vaccination on binding to the four epitopes identified above (Figure 4.3). For samples in the Moderna Trial Cohort, there was significantly decreased binding to the CTD epitope and L-HR2 epitope ($p=0.008$, $p=0.011$, Wilcoxon rank-sum test with Bonferroni correction) at the later timepoint post first dose (day 119) compared to the earlier timepoint (day 36) (Figure 4.3A). To examine the effect of dosage, we compared 100 ug and 250 ug mRNA-1273 groups for those between the age of 18 to 55, as that was the only age group included for the 250 ug dose (Figure 4.3B). There was no significant difference by vaccine dosage for any of the four epitope regions (NTD, CTD, FP, or L-HR2). Participant age was also examined as a variable; there appeared to be a difference in epitope binding in the L-HR2 region, but this did not survive multiple testing correction (Figure 4.3C).

In infected individuals, the effect of time post symptom onset on epitope binding was examined using non-hospitalized infected individuals in the HAARVI Cohort, who were sampled between 26 and 309 days post symptom onset (Supplemental Figure 4.2A). Samples were binned into three groups: 0-60, 60-180, and 180-360 days post symptom onset. At all times post symptom onset there was no significant difference in binding to the four identified epitopes ($p>0.05$, M.W.W.). Individuals in the HAARVI Cohort were given either the Moderna mRNA-1273 or Pfizer/BioNTech BNT162b2 mRNA vaccine, and comparison of the epitope binding response between the two vaccine types revealed no significant differences in all epitope regions (Supplemental Figure 4.2B, $p>0.05$, M.W.W.).

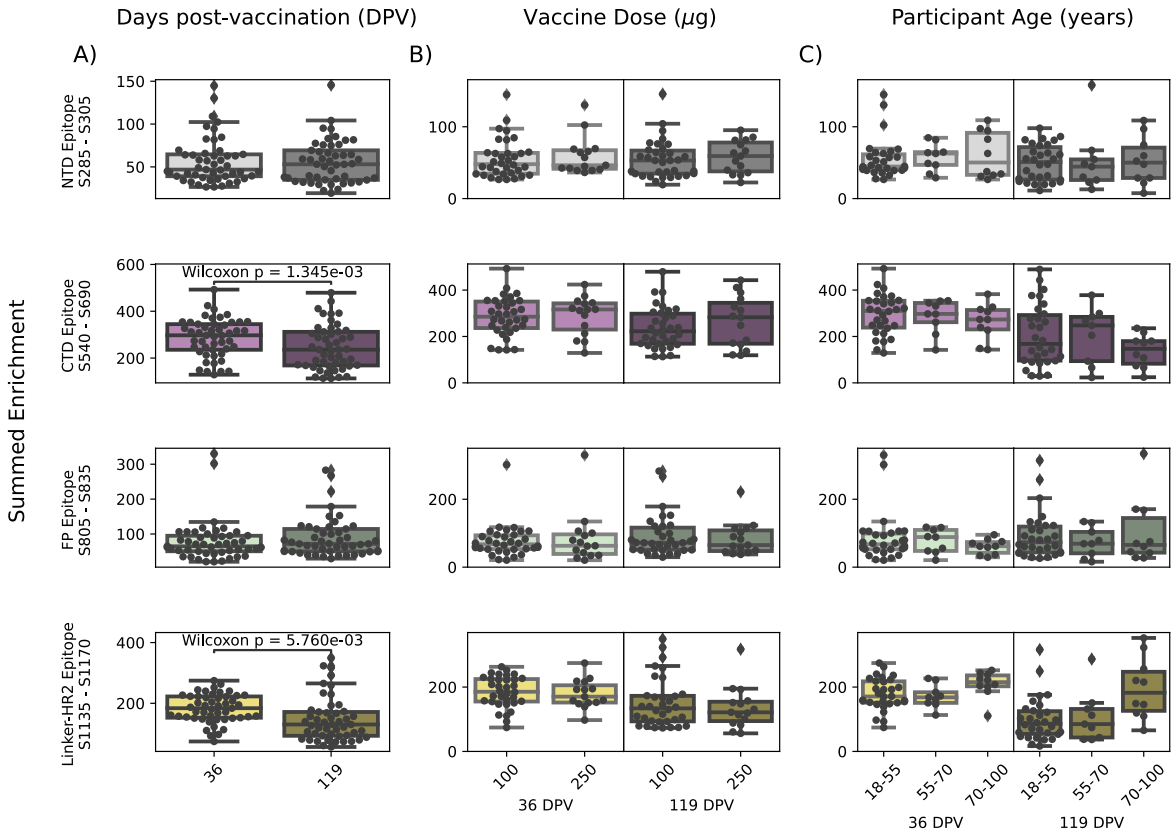


FIGURE 4.3: Comparison of epitope binding for NIH Moderna Trial subgroups. Boxplots of summed wild-type enrichment within epitope binding regions for samples grouped by (A) timepoint post vaccination, (B) vaccine dose, or (C) participant age. Samples were taken at either at 36 (n=64) or 119 (n=64) days post vaccination. (B) and (C) are additionally separated by timepoint post vaccination. Results of a Wilcoxon rank-sum test between the groups appears only where $p < 0.05$ after Bonferroni multiple testing correction (36 group comparisons). Figures containing all p-values for both replicate batches are available at https://github.com/matsengrp/phage-dms-vacc-analysis.

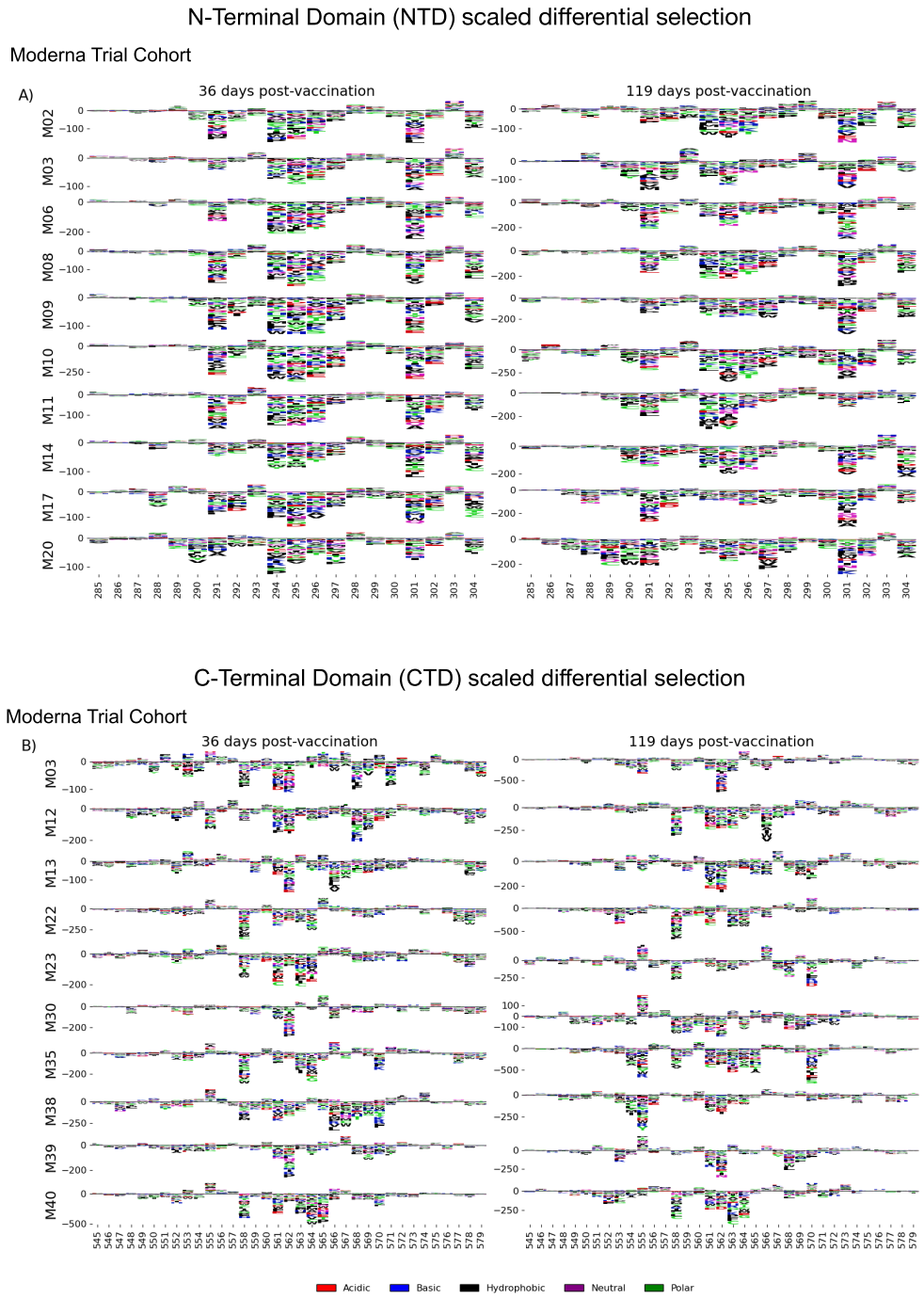
Infection and vaccination shape pathways of escape

The Spike Phage-DMS library contains peptides with every possible single amino acid substitution in addition to the wild-type sequence, enabling us to assay the impact of mutations on antibody binding. The effect of site-specific substitutions in critical antibody binding regions not only provides a high-resolution picture of the likely epitope intervals, but also identifies mutations that confer escape within the binding region. The effect of each mutation on serum antibody binding was quantified by calculating its scaled differential selection value, a metric that reports log fold change of mutant binding affinity over wild-type binding affinity at any given

site (see Methods)[119]. Site mutations that cause a loss of binding when compared to the wild-type peptide centered at that same site are reported as having negative differential selection values, whereas those that bind better than the wild-type peptide have positive differential selection values. In order for differential selection to be meaningful, however, we must ensure that we do not include weak or sporadic signals that may be due to non-specific binding. Accordingly, we set a threshold of summed wild-type peptide binding in any one region. By doing so, we lose samples in the analysis but can be confident in the results presented by samples passing this curation step (Supplemental Figure 4.3). For samples that passed this threshold, we compared the effect of prior infection and/or time post vaccination on the pathways of escape in each epitope region as follows. Plots depicting the effect of mutations for all samples are publicly explorable at {URL redacted until publication}.

N-Terminal Domain (NTD) and C-terminal Domain (CTD)

We examined the sites of escape within the NTD and CTD epitope regions, focusing on vaccinated individuals from the Moderna Trial Cohort because these epitopes were notable targets of the vaccine response and not commonly found in infected individuals. Vaccination elicited antibodies with a strikingly uniform escape profile in the NTD epitope across samples (Figure 4.4A), with the majority of samples being sensitive to mutation at sites 291, 294-297, 300-302, and 304, which are in the very C terminal portion of NTD as well as the region between NTD and RBD. The CTD region appeared to consist of multiple epitopes, the dominant being at position 545 to 580 (termed CTD-I). Antibodies that bound to this dominant CTD epitope had a less uniform escape profile, but sites 561 and 562 were common sites of escape in most samples (Figure 4.4B). For antibodies to both the NTD and CTD-I epitopes, the pathways of escape tended to drift over time and were different at 119 days post vaccination as compared to 36 days post vaccination.



<u>FIGURE 4.4:</u> NTD and CTD-1 epitope escape profiles. (A and B) Logo plots depicting the effect of mutations on epitope binding in either the NTD (A) or CTD-1 (B) epitope for paired samples from the Moderna Trial Cohort. The height of the letters corresponds to the magnitude of the effect of that mutation on epitope binding, i.e. its scaled differential selection value. Letters below zero indicate mutations that cause poorer antibody binding as compared to wild-type peptide, and letters above zero indicate mutations that bind better than the wild-type peptide. Letter colors denote the chemical property of the amino acids. Logo plots on the left and right are paired samples from the same individual, with the participant ID noted on the left.

Fusion Peptide (FP)

Antibodies against the FP epitope are strongly stimulated after infection but are less strongly induced after subsequent vaccination (Figure 4.2). Thus, we investigated whether the pathways of escape for serum antibodies also changed after vaccination within samples from previously infected individuals in the HAARVI Cohort. The escape profiles of antibodies in paired samples that strongly bound to the FP epitope both after infection and after subsequent vaccination are shown as a logo plot (Figure 4.5A). The major sites of escape within the FP epitope for these samples were sites 819, 820, 822, and 823, and these sites of escape did not appear to change after vaccination although we noted there was more variability in the escape profiles after vaccination.

We next examined the pathways of escape for FP binding antibodies in vaccinated individuals from the Moderna Trial Cohort. In people with no prior infection, vaccination induced diverse pathways of escape in the FP region (Figure 4.5B). For example, for participant M10 escape was focused on sites 814, 816, and 818, whereas for participant M38 escape was focused on 819, 820, and 823. There appeared to be some differences in the escape profile at 119 days as compared to 36 days post vaccination, as exemplified by participants M15, M17, and M20. However, in general many of the major sites of escape were shared at both timepoints within each individual and as a group.

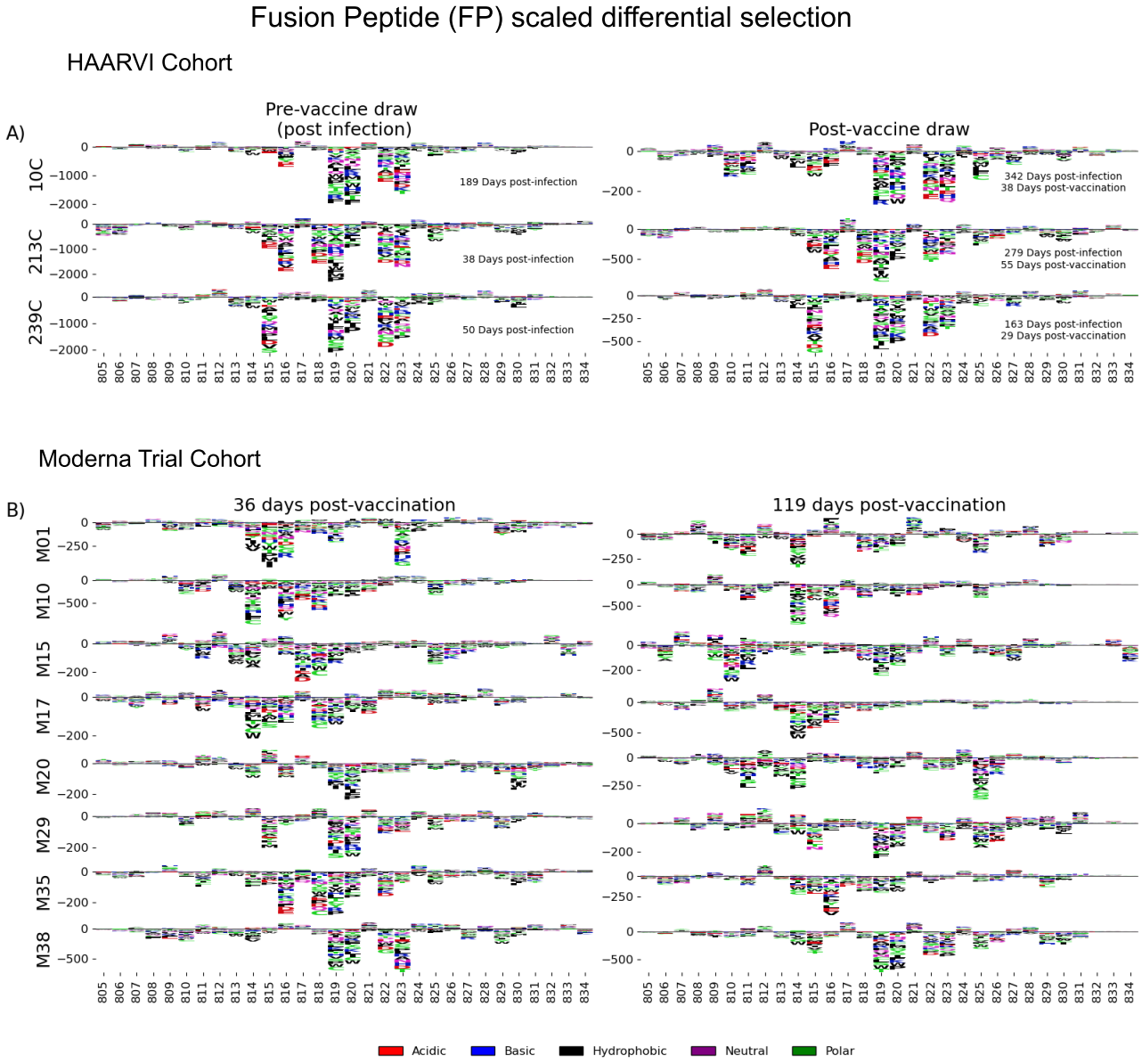


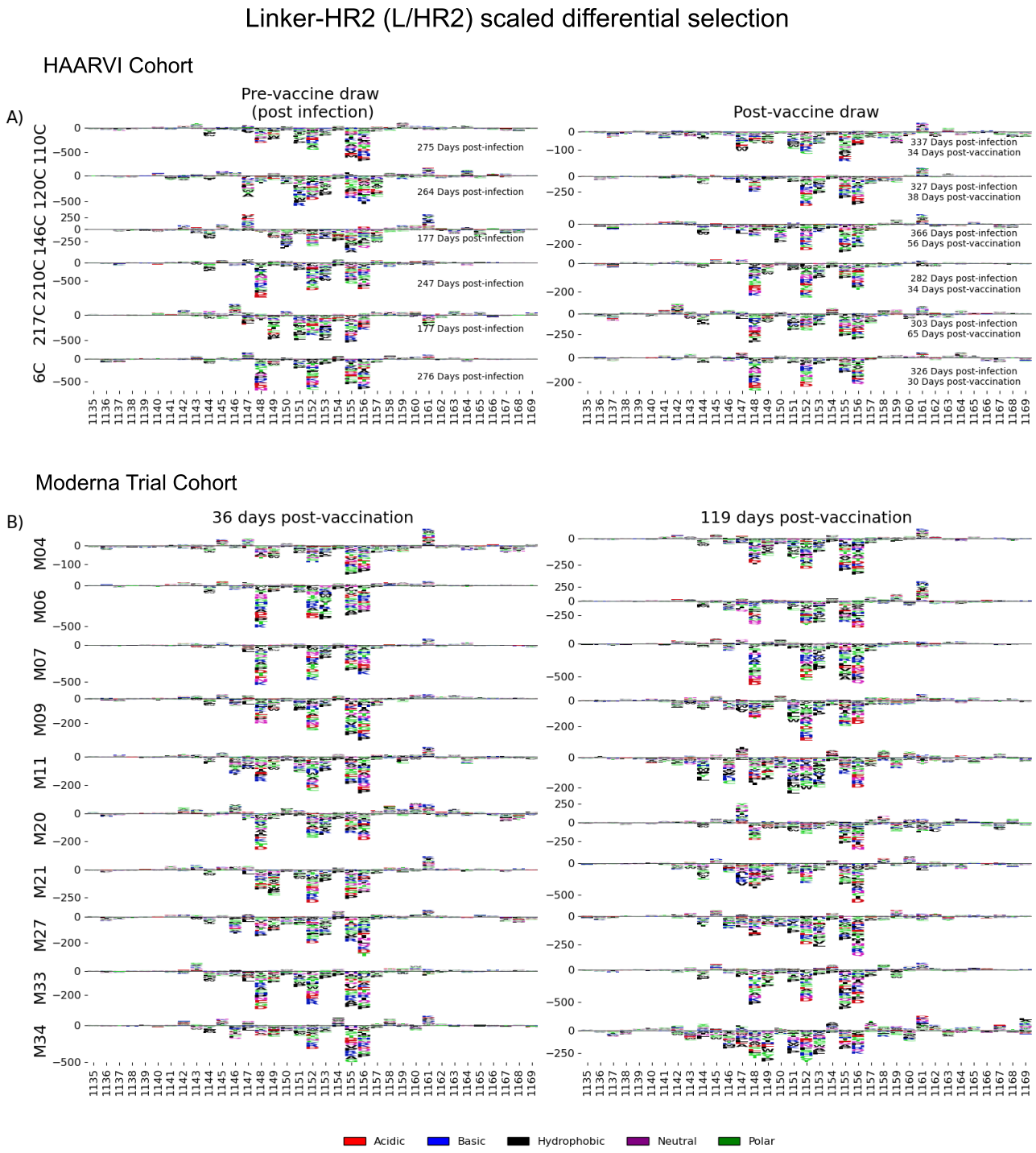
FIGURE 4.5: FP epitope escape profiles. (A and B) Logo plots depicting the effect of mutations on epitope binding within the FP epitope region for paired samples from the (A) HAARVI Cohort or (B) Moderna Trial Cohort. Details are as described in Fig 4.4.

Linker-Heptad Repeat 2 (L-HR2)

In order to determine the effect of prior infection on the binding profiles of antibodies after vaccination within the L-HR2 epitope region, we explored the pathways of escape for paired samples from patients with prior infection in the HAARVI Cohort before and after

vaccination as was done for the FP region. Samples from previously infected individuals with no vaccination history had diverse pathways of escape within the L-HR2 epitope (Figure 4.6A). For example, site 1149 was only sensitive to mutation for participant 217C, and site 1157 was only sensitive to mutations for participants 120C and 146C. In contrast, the samples from vaccinated individuals, regardless of infection history, tended to have a uniform pathway of escape. The most prominent and consistent sites of escape for vaccinated individuals, both with and without prior infection, were at sites 1148, 1152, 1155 and 1156. Of note, the pre-vaccination sample from an individual with prior infection requiring hospitalization (participant 6C) displayed an escape profile highly similar to those from vaccinated individuals, and this escape profile did not change after vaccination.

To see whether the pathways of escape changed over time after vaccination, we visualized the escape mutations within the L-HR2 epitope for the samples in the Moderna Trial Cohort at 36 and 119 days after the first dose of vaccine (Figure 4.6B). We saw a highly uniform pattern of escape for most samples at day 36 and 119, again with escape mainly occurring at sites 1148, 1152, 1155 and 1156. For some participants, such as M11, M34, and M35, the escape mutations appeared to drift over time, but the major sites of escape remained the same.



<u>FIGURE 4.6:</u> L-HR2 epitope escape profiles. (A and B) Logo plots depicting the effect of mutations on epitope binding within the L-HR2 epitope region for paired samples from the (A) HAARVI Cohort or (B) Moderna Trial Cohort. Details are as described in Fig 4.4.

4.4 Discussion

In this study, we comprehensively profiled the antibody response to the SARS-CoV-2 Spike protein, including pathways of escape from sera in individuals with diverse infection and vaccination histories. We identified four major targets of antibody responses outside of the core RBD domains, in the NTD, CTD, FP, and Linker-HR2 regions. Vaccinated individuals as well as individuals with severe infection requiring hospitalization both had antibodies to these four epitope regions, whereas individuals with mild infection that did not require hospitalization preferentially targeted only FP and L-HR2. In previously infected cases, the epitope binding patterns changed over time after vaccination, with decreased binding to both the CTD and L-HR2 epitopes. However, there was not uniform decay across all four epitopes, indicating that waning antibody titers may not occur for all epitopes equally. Other factors such as vaccine dose (100 ug or 250 ug), vaccine type (BNT162b2 or mRNA-1273), and participant age did not significantly affect the specificity of the antibody response.

We explored the pathways of escape for antibodies binding to these key regions in infected and vaccinated people. We defined for the first time the escape pathways for NTD and CTD-I binding antibodies, epitopes that were commonly found in vaccinated individuals but not in infected individuals. In the case of the NTD epitope, which was located at the C-terminal end of the NTD, escape mutations were uniform and consistent amongst vaccinees, while pathways of escape were more diverse for CTD-I antibodies. Individuals with antibodies that strongly bound the FP epitope had focused escape profiles, with the majority of escape occurring at sites 819, 820, 822, and 823, although the sample size of this group is small (N=3). Vaccination did not greatly alter the escape profile in previously infected individuals, nor did vaccination alone induce a strong or uniform response at the FP epitope. In contrast, antibodies that bind in the L-HR2 epitope region after infection have diverse pathways of escape, while after vaccination they appear to converge on a more uniform pathway of escape that includes mutations at sites 1148, 1152, 1155 and 1156. Interestingly, these are also the sites of contact for a cross-reactive HR2-specific antibody isolated from a mouse sequentially immunized with the MERS and SARS Spike proteins[160]. This hints that a singular antibody clonotype could be elicited when exposed to a stabilized Spike protein, dominating the response in the L-HR2 region.

We also observed some drift in the pathways of escape within a single person over time after vaccination. This mirrors findings from a recent study that examined sites of escape for RBD-specific antibodies in serum samples from the same Moderna Trial Cohort as used in this study[62]. Together these results suggest that the B cell response after vaccination with Spike

mRNA continues to evolve over time. Multiple studies have demonstrated that SARS-CoV-2-specific B cells undergo continued somatic hypermutation in the months after infection, likely due to antigen persistence[161, 162]. Spike antigen has been detected in the lymph nodes at least 3 months after vaccination with BNT162b2, and continued maturation of germinal center B cells could be a possible explanation for the changes in epitope binding we observed[163]. Alternatively, turnover of short-term plasma cells and memory B cells could account for loss of antibody binding to certain epitopes.

Our study has important limitations worth noting. Because the Spike Phage-DMS library displays 31 amino acid peptides, we are unable to detect antibodies that bind to conformational epitopes and/or glycosylated epitopes. This is demonstrated by the lack of observable binding to the RBD region, a domain with complex folding and known target of antibodies from infected and vaccinated individuals. However, prior studies of RBD epitopes have already been reported using an overlapping set of samples from the HAARVI Cohort, however, and together these results paint a more complete picture of epitopes across the Spike protein[61, 62]. Finally, we only have 5 individuals within the hospitalized group and this small sample size limits our ability to make conclusions about epitope binding in those with severe infection.

Our finding that vaccinated individuals have a broader response across the Spike protein than infected individuals may have important implications for immune durability against future SARS-CoV-2 variants. Evidence suggests that a polyclonal antibody response that is resistant in the face of multiple mutations is necessary for long-lasting immunity against a mutating viral pathogen[164]. Thus, the polyclonal response to vaccination may provide greater protection from infection than the more focused response after infection. However, the number of epitopes targeted provides just one benchmark and the ability to escape at the population level could also be influenced by the diversity of individuals' antibody responses at each epitope and thus the likelihood that a single escape mutation could be widely selected. At one S2-domain epitope region (L-HR2) vaccination induced uniform sites of escape that may be due to a singular type of antibody that would allow escape by the same mutations for all vaccinated people. However, epitopes in the S2 domain tend to be in highly conserved regions with important functions that constrain the virus' ability to mutate, making escape from these antibodies less likely than for RBD, where escape is already common. Indeed, mutations in the FP and L-HR2 epitopes are not arising in the global population of SARS-CoV-2[157], providing some suggestion that these regions may be constrained[97, 165]. Overall, further studies of the functional capacity of these vaccine-elicited antibodies targeting epitopes outside of RBD are warranted, to provide a path

towards a polyclonal response to epitopes across the full Spike protein. This comprehensive view may further the goal of a more universal coronavirus vaccine that eliminates the need for continual updates of the SARS-CoV-2 vaccine strain due to mutations in variable regions on Spike.

4.5 Methods

Sample collection

Moderna Trial Cohort

We obtained post-vaccination serum samples via the National Institute of Allergy and Infection Disease that were taken as part of a phase I clinical trial testing the safety and efficacy of the Moderna mRNA-1273 vaccine (NCT04283461)[159]. All samples were de-identified and thus all work was approved by the Fred Hutchinson Cancer Research Center Institutional Review Board as nonhuman subjects research. Trial participants were given either 100 ug or 250 ug doses of the mRNA-1273 vaccine, and serum was sampled from all trial participants at 36 days and 119 days post vaccination. See Table 4.1 for detailed metadata related to each participant and serum sample.

HAARVI Cohort

We obtained plasma samples from individuals enrolled in the Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) study conducted in Seattle[157]. Individuals were either enrolled upon PCR confirmed diagnosis with SARS-CoV-2 infection or as control subjects prior to receiving vaccination with either BNT162b2 (Pfizer/BioNTech) or mRNA-1273 (Moderna). See Table 4.1 for detailed metadata related to each participant and plasma sample. For convenience, all plasma and serum samples in this study are referred to as serum. This study was approved by the University of Washington Institutional Review Board.

Spike Phage-DMS assay

The Spike Phage-DMS library used in this study contained 24,820 designed peptides that tile across the length of the Spike protein. Peptides are each 31 amino acids long and tile

by 1 amino acid increments, and correspond to either the wild-type sequence or a sequence containing a single mutation. Serum samples were profiled using the Spike Phage-DMS library as previously described[157]. Following this method, the Spike Phage-DMS library was diluted in Phage Extraction Buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 6 mM $MgSO_4$) to a concentration of 2.964 x 10^9 plaque forming units/mL, which corresponds to approximately 200,000-fold coverage of each peptide. 10 uL of serum or plasma was added to 1 mL of the diluted library and incubated in a deep 96-well plate overnight at 4ºC on a rotator. 40uL of a 1:1 mixture of Protein A and Protein G Dynabeads (Invitrogen) were added to each well and then incubated at 4ºC for 4 hours on a rotator. Beads bound to the antibody-phage complex were magnetically separated and washed 3x with 400 uL wash buffer (150 mM NaCl, 50 mM Tris-HCl, 0.1% [vol/vol] NP-40, pH 7.5). Beads were resuspended in 40 uL of water and lysed at 95ºC for 10 minutes. The diluted Spike Phage-DMS library was also lysed to capture the starting frequencies of peptides. All samples were run twice, once each with two independently generated Spike Phage-DMS libraries.

DNA from lysed samples were amplified and sequenced as previously described[157]. Two rounds of PCR were performed using Q5 High-Fidelity 2X Master Mix (NEB). For the first round of PCR, 10uL of lysed phage was used as the template in a 25 uL reaction using primers described in [157]. For the second round of PCR, 2 uL of the round 1 PCR product was then used as the template in a 50 uL reaction, with primers that add dual indexing sequences on either side of the insert. PCR products were then cleaned using AMPure XP beads (Beckman Coulter) and eluted in 50 uL water. DNA concentrations were quantified via Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). Equimolar amounts of DNA from the samples, along with 10X the amount of the input library samples, was pooled, gel purified, and the final library was quantified using the KAPA Library Quantification Kit (Roche). Pools were sequenced on an Illumina HiSeq 2500 machine using the rapid run setting with single end reads.

Sample curation and replicate structure

All sample IP's and downstream analysis were run in duplicate across two separate phage display library batches to ensure reproducibility, with the exception of the four acute samples from hospitalized HAARVI participants which were run in singlicate. All results were cross checked with the set of batch replicates to ensure significance fell within one order of

magnitude where applicable. For brevity, we present only figures resulting from the single complete set of batch-specific replicates, however, all figures using the second set of library batch replicates are available (see Code and Data Availability). Additionally, some samples were run with "in-line" technical replicates within the same batch. In the case with more than one technical replicate, we selected the sample with the highest reads mapped from each set of batch replicates for our downstream analysis.

Short read alignment and peptide counts processing

Samples were aliquoted and sequenced targeting 10X coverage of total sample reads to the peptide library reference. We demultiplexed samples using Illumina MiSeq Reporter software. Post sample demultiplexing, we used a *Nextflow* pipeline to process the peptide counts as well as alignment stats for all samples[166]. The tools and parameters describing the workflow are as follows. The index creation and short-read alignment step were done using *Bowtie2.* During alignment we allowed for zero mismatches in the default seed length of each read (20, very sensitive) after trimming 32 bases from the 3' end of each 125bp read to match the 93 bp peptides in our reference library[167]. *Samtools* was subsequently used to gather sequencing statistics as well as produce the final peptide counts using the *stats* and *idxstats* modules. Finally, the pipeline collected all reference peptide alignment counts and merges them into a single *xarray* dataset coupling sample and peptide metadata with their respective count. Alignment stats for all replicates are seen in Supplementary Figure 4.6.

Each of the processing steps described here, as well as downstream analysis and plotting, were run using static and freely available Docker containers for reproducibility. We provide an automated workflow and the configuration scripts defining exact parameters. See Code and Data Availability section for more information.

Epitope binding region identification

Principal Component Analysis (PCA) via Singular Value Decomposition (SVD) was performed on each set of batch replicates using the scikit-learn package[168]. We first subset our dataset to only include wildtype peptide count enrichments from either infected or

vaccinated individuals as input. This curation resulted in the matrix, X of size $n \times p$ with n biologically distinct replicates and p enrichment features across the spike protein. All enrichment values were calculated as a fold change in the frequency for any one sample enrichment over the library control enrichment at the same sites. Each feature was mean centered before performing the PCA such that the covariance matrix of X is equivalent to $X^T X/(n-1)$. We can then use the eigendecomposition, $X = USV^T$, to describe the data. The principal axes in feature space are then represented by the columns of V and represent the direction of maximum variance in the data. Supplemental Figure 4.1 shows three facets of this decomposition; Supplemental Figure 4.1A: the unit scaled sample "scores" represented by the columns to visualize sample relationship in principal component space; Supplemental Figure 4.1B: component loadings (scaled by the square root of the respective eigenvalues in S); and Supplemental Figure 4.1C: the first three principal axes/directions in feature space plotted as a function of the WT peptide feature location on the Spike protein. Together, these provide a visualization of key features in the data used in our downstream analysis. We chose our epitope regions as contiguous regions of nonzero value in the loadings in the first three principal axes.

Identifying high-resolution pathways of escape

In order to ensure reliable measurements of differential selection of single AA variants compared to the ancestral sequence variant, we threw out samples whose respective sum of wild-type enrichment was below a threshold set for each of the defined binding regions (Supplemental Figure 4.4). Once curated, we computed the log-fold change in each of the 19 possible variant substitutions at each site. This metric was then scaled by the average of the wild-type sequence enrichment coupled with both the preceding and following wild-type peptide enrichments at any given site. To evaluate escape at each site, we then sum the differential selection metric as described for each variant at a site to examine a more complete picture of the data defining escape patterns in each sample group.

CODE AND DATA AVAILABILITY

We provide a fully reproducible automated workflow which ingests raw sequencing data and performs all analyses presented in the paper. The workflow defines and runs the processing steps within publicly available and static Docker software containers, including *phippery* and *phip-flow* described in the Methods section. The source code, Nextflow script, software dependencies, and instructions for re-running the analysis can be found at {URL redacted until publication}.

The generalized PhIP-Seq alignment and count generation pipeline script can be found at https://github.com/matsengrp/phip-flow. A template and documentation for the alignment pipeline configuration is available at https://github.com/matsengrp/phip-flow-template. Finally, we provide a python API, *phippery,* to query the resulting dataset post-alignment that can be found at https://github.com/matsengrp/phippery.

All raw sequencing data was submitted to the NCBI SRA. Pre-processed enrichment data is available upon request. Additionally, differential selection data and more can be explored interactively using the dms-view toolkit available at {URL redacted until publication}.

For more information, please email jgallowa@fredhutch.org.

STATISTICAL ANALYSIS

Estimates of significance presented between group continuous distributions of wild-type enrichment were reported using a *Mann-Whitney Wilcoxon* test with the exception of analysis that included only paired longitudinal samples - such as the comparison of 36-and-119 Days post-vaccination - in this case we used a *Wilcoxon signed-rank* test. Bonferroni correction was applied where applicable and adjusted P values < 0.05 were presented as significant. All statistical analysis were done using *Python 3.6* and plotted using the *statannot* package found here https://github.com/webermarcolivier/statannot. The static Docker container used for all statistical analysis is publicly hosted at https://quay.io/repository/matsengrp/vacc-ms-analysis.

4.6 Notes

ACKNOWLEDGEMENTS

We thank Kevin Sung, Thayer Fisher, and Noah Simon for providing advice regarding computational analyses. We thank Cassie Sather and the Genomics core facility for assistance with sequencing. We thank Laura Jackson (Kaiser Permanente), Chris Roberts, Catherine Luke, and Rebecca Lampley [National Institute of Allergy and Infectious Diseases (NIAID), NIH] for assistance with obtaining the mRNA-1273 phase 1 trial vaccine samples. We also thank all research participants and study staff of the Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) study. This work was supported by NIH grants AI138709 (PI Overbaugh) and AI146028 (PI Matsen). Julie Overbaugh received support as the Endowed Chair for Graduate Education (FHCRC). The research of Frederick Matsen was supported in part by a Faculty Scholar grant from the Howard Hughes Medical Institute and the Simons Foundation. Scientific Computing Infrastructure at Fred Hutch was funded by ORIP grant S10OD028685.

CONTRIBUTIONS

J.O. and F.A.M. conceived the project; M.E.G., J.G., F.A.M., and J.O. led the design of the study; H.Y.C. led the HAARVI study, with C.R.W., J.K.L., and N.F. involved in sample collection. M.E.G. performed all experiments, and J.G. performed all computational and data analyses with F.A.M. advising. M.E.G., J.G., F.A.M, and J.O. wrote the paper with input from all authors.

ETHICS DECLARATIONS

Competing Interests

H.Y.C. reported consulting with Ellume, Pfizer, The Bill and Melinda Gates Foundation, Glaxo Smith Kline, and Merck. She has received research funding from Gates Ventures, Sanofi Pasteur, and support and reagents from Ellume and Cepheid outside of the submitted work.

Chapter 5

ONGOING AND FUTURE DIRECTIONS

Methods and technologies all originally developed to tackle the HIV-1 pandemic laid the groundwork for the rapid response made at the outset of the SARS-CoV-2 pandemic in late 2019 and 2020, and the studies in this thesis exemplify this concept. By building a system that could rapidly map fine antibody epitopes for HIV-1, we were able to quickly pivot this system to be used with SARS-CoV-2 mere months after the virus was sequenced. And similarly, less than a year after vaccines were deployed against SARS-CoV-2, we were able to map differences in the antibody response between infected and vaccinated individuals in detail.

Given our lab's access to samples from established HIV cohorts in Kenya and the HAARVI cohort for SARS-CoV-2 infections in Seattle, there are a plethora of scientific questions that could be answered using Phage-DMS regarding critical antibody responses and pathways of escape. Below I describe ongoing work and potential future directions for studying virus-specific antibody responses.

5.1 Applying Phage-DMS to HIV-1 antibodies

As described in section 1.2.2, antibodies with non-neutralizing effector functions have been found to be correlated with protection in a partially successful HIV vaccine trial and have been associated with positive outcomes in infants infected with HIV-1 [36, 39]. Studies from our lab have found that ADCC-mediating IgG in the breast milk of HIV positive pregnant women in Kenya correlated with reduced mother-to-child transmission (MTCT), and serum ADCC activity correlated with a trend towards reduced MTCT and improved clinical outcome in infected infants [38, 39, 138]. Therefore, we hypothesized that passively transferred antibodies from non-transmitting mothers potentially protected their infants from infection. Early in my PhD I took steps towards identifying mothers with high serum ADCC activity and screening cultured B cell supernatants for binding activity, and others in the lab have taken on the task of isolating monoclonal antibodies from these non-transmitting mothers. Phage-DMS can be used to aid in the often arduous process of identifying the epitopes of new antibodies that have been isolated. Protective antibodies isolated and mapped in this way can provide a blueprint for building an effective HIV-1 vaccine antigen.

5.2 Characterizing novel SARS-CoV-2 monoclonal antibodies

In the studies described in Chapter 3 and 4 we found that there were antibodies from individuals with prior COVID-19 infection that tended to bind to an epitope that overlaps the fusion peptide region as well as the stem helix just upstream of heptad repeat 2, both in the more conserved S2 subunit. Prior studies have shown that antibodies targeting these epitopes can be neutralizing, and thus far one group has isolated a broadly neutralizing mAb targeting the stem helix from a SARS-CoV-2 infected individual [110, 111, 169]. The fusion peptide of other class I fusion proteins such as HIV Env and influenza hemagglutinin is a known target of neutralizing antibodies, and recently SARS-CoV-2 fusion peptide binding antibodies have been identified [170]. Indeed, the fusion peptide specific antibody VRC34.01 has driven a new rationally designed HIV vaccine antigen [171]. The relative conservation of these epitope regions on the S2 subunit of SARS-CoV-2, as opposed to the more mutationally tolerant RBD and NTD domains, makes them an attractive target for vaccination. Therefore, we set out to isolate S2-specific antibodies from a recovered COVID-19 patient.

Thus far we have sorted single S2-specific B cells from blood samples taken at 34 days post symptom onset from a single individual (20C) enrolled in the HAARVI cohort [157]. This individual was chosen because we saw serum binding to both the fusion peptide and stem helix epitopes by Phage-DMS. We are in the process of cloning and characterizing the corresponding antibodies, but preliminary results indicate that these antibodies bind to epitope present on either the S2 subunit alone, or present on the Spike trimer as well as the S2 protein. As shown in Figure 5.1, we have identified six antibodies that have a range of binding to either the S2 monomer or Spike trimer by ELISA. Although we have yet to identify a neutralizing antibody, more antibodies are currently being cloned and we have plans to test all antibodies for non-neutralizing activity. All isolated antibodies will be screened with the Spike Phage-DMS library to see if they bind to epitopes previously identified using polyclonal serum.

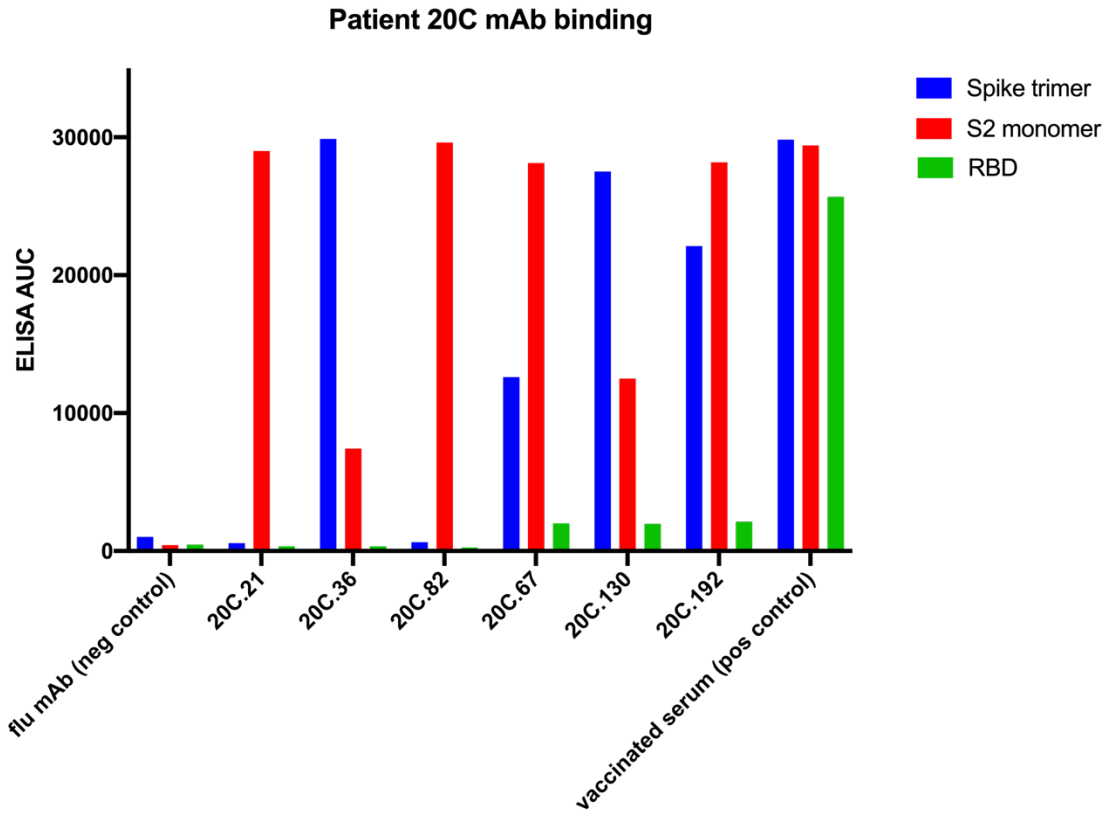


Figure 5.1. Binding of monoclonal antibodies isolated from patient 20C to Spike antigens. We measured binding by ELISA for six monoclonal antibodies isolated from patient 20C, as well as a flu-specific negative control antibody (Fi6v3) and a positive control serum sample from a vaccinated individual. Proteins were coated at 2 ug/mL, and antibodies were tested at six dilutions starting at 10 ug/mL and diluted 1:4. Serum concentration started at 1:50 and then 1:4 dilutions were prepared from there. Activity in background wells was subtracted, and the area under the curve is reported. Results are from one replicate.

We also have plans to isolate antibodies from vaccinated individuals, as we have seen that epitopes targeted by infected and vaccinated individuals differ (Chapter 4). Specifically, we found that vaccination alone did not strongly elicit fusion peptide binding antibodies and that after vaccination there was a uniform antibody response to the stem helix epitope. By isolating mAbs from vaccinated individuals as well as infected individuals, we would be able to compare and contrast the functions of these antibodies as well as the mutations that could lead to escape. Given that vaccines elicit neutralizing antibodies that are even more RBD-focused than in infection [62], it will be important to understand the role of non-RBD antibodies and whether they could influence immunity against new variant strains of SARS-CoV-2. The longer term goal of this work would be to identify protective antibodies that are difficult to escape by viral

evolution in order to guide the design of the next generation of cross-protective coronavirus vaccines.

5.3 Improvements to Phage-DMS

In the future there may be ways to update and improve Phage-DMS. One likely step will be to update the Spike Phage-DMS library as new variants arise. The introduction of new mutations can potentially alter antibody binding to nearby epitopes in what is known as an epistatic interaction. An example of this can be seen in Chapter 3, where we demonstrated that in some cases serum binding to mutant peptides spanning site 614 was changed when site 614 was mutated from a D to a G (Figure 3.6). Therefore, as new strains arise with mutations in the Spike protein, it will likely be necessary to create new phage displaying peptides that incorporate these mutations and add them to the Spike Phage-DMS library.

Phage-DMS can provide information about the sensitivity of antibody binding to single amino acid mutations, and in Chapter 2 we showed that the effect of mutations on monoclonal antibody binding was relatively quantitative. However, when working with antibodies in polyclonal serum it would be helpful to understand not just the relative binding to different epitopes, but the absolute amount of antibodies directed to each epitope. One way to achieve this information could be to include a set of control peptides in the phage library and to spike in a known amount of monoclonal antibody to compare against. This would also aid in normalizing and comparing results across Phage-DMS experiments performed on different days. Providing a quantitative measure of antibody binding would allow us to get the maximum amount of information from each Phage-DMS experiment and could be integrated into our ever-improving analysis pipeline.

Overall we envision this system being applied to not only viral immunology, but beyond. Phage-DMS has the potential to map binding sites not just between antibodies and other viral entry proteins, but any protein-protein interaction of interest. As DNA synthesis technology improves, longer peptide sequences can be designed, potentially allowing for more complex epitopes to be expressed on the surface of phage. The ease of use, scalability, and relative cost effectiveness of Phage-DMS make it an attractive tool to apply in many different circumstances.

5.3 Conclusion

In this thesis I have described not only a new method of mapping antibody epitopes (Chapter 2), but also used this method to provide insight into the epitopes and pathways of escape for antibodies derived from SARS-CoV-2 infection or vaccination (Chapters 3 and 4). While it may feel otherwise, relatively little time has passed since the beginning of the SARS-CoV-2 pandemic. Thus, it is startling how much has already been uncovered about the biology of this virus and the immune response to it, while at the same time there remain many unanswered questions. I look forward to extending these studies to better understand immunity to pandemic viruses while also laying groundwork to prepare for the next one.

BIBLIOGRAPHY

1. UNAIDS. *Global HIV & AIDS statistics — Fact sheet*. 2021 [cited 2021 8/15/21]; Available from: https://www.unaids.org/en/resources/fact-sheet

.

2. WHO. *WHO Coronavirus (COVID-19) Dashboard*. [cited 2021 8/15/21]; Available from: https://covid19.who.int/.

3. White, J.M., et al., *Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme.* Crit Rev Biochem Mol Biol, 2008. **43**(3): p. 189-219.

4. Apellániz, B., et al., *The three lives of viral fusion peptides.* Chem Phys Lipids, 2014. **181**: p. 40-55.

5. Checkley, M.A., B.G. Luttge, and E.O. Freed, *HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation.* J Mol Biol, 2011. **410**(4): p. 582-608.

6. Berger, E.A., P.M. Murphy, and J.M. Farber, *Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease.* Annu Rev Immunol, 1999. **17**: p. 657-700.

7. Frost, S.D., et al., *Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection.* Proc Natl Acad Sci U S A, 2005. **102**(51): p. 18514-9.

8. Gray, E.S., et al., *Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection.* J Virol, 2007. **81**(12): p. 6187-96.

9. Wei, X., et al., *Antibody neutralization and escape by HIV-1.* Nature, 2003. **422**(6929): p. 307-12.

10. Prabakaran, P., et al., *Structure and function of the HIV envelope glycoprotein as entry mediator, vaccine immunogen, and target for inhibitors.* Adv Pharmacol, 2007. **55**: p. 33-97.

11. Huang, Y., et al., *Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19.* Acta Pharmacol Sin, 2020. **41**(9): p. 1141-1149.

12. Shang, J., et al., *Cell entry mechanisms of SARS-CoV-2.* Proc Natl Acad Sci U S A, 2020. **117**(21): p. 11727-11734.

13. Daly, J.L., et al., *Neuropilin-1 is a host factor for SARS-CoV-2 infection.* Science, 2020. **370**(6518): p. 861-865.

14. Cantuti-Castelvetri, L., et al., *Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity.* Science, 2020. **370**(6518): p. 856-860.

15. Wei, C., et al., *HDL-scavenger receptor B type 1 facilitates SARS-CoV-2 entry.* Nat Metab, 2020. **2**(12): p. 1391-1400.

16. Wang, K., et al., *CD147-spike protein is a novel route for SARS-CoV-2 infection to host cells.* Signal Transduct Target Ther, 2020. **5**(1): p. 283.

17. Chen, R.E., et al., *Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies.* Nat Med, 2021. **27**(4): p. 717-726.

18. Harvey, W.T., et al., *SARS-CoV-2 variants, spike mutations and immune escape.* Nat Rev Microbiol, 2021. **19**(7): p. 409-424.

19. Pallesen, J., et al., *Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen.* Proc Natl Acad Sci U S A, 2017. **114**(35): p. E7348-e7357.

20. Corbett, K.S., et al., *SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness.* Nature, 2020. **586**(7830): p. 567-571.

21. Hsieh, C.L., et al., *Structure-based design of prefusion-stabilized SARS-CoV-2 spikes.* Science, 2020. **369**(6510): p. 1501-1505.

22. Forthal, D.N., *Functions of Antibodies.* Microbiol Spectr, 2014. **2**(4): p. 1-17.

23. Ruprecht, C.R., et al., *MPER-specific antibodies induce gp120 shedding and irreversibly neutralize HIV-1.* J Exp Med, 2011. **208**(3): p. 439-54.

24. de Rosny, E., et al., *Binding of the 2F5 monoclonal antibody to native and fusion-intermediate forms of human immunodeficiency virus type 1 gp41: implications for fusion-inducing conformational changes.* J Virol, 2004. **78**(5): p. 2627-31.

25. Lorizate, M., et al., *Recognition and blocking of HIV-1 gp41 pre-transmembrane sequence by monoclonal 4E10 antibody in a Raft-like membrane environment.* J Biol Chem, 2006. **281**(51): p. 39598-606.

26. Excler, J.L., et al., *Nonneutralizing functional antibodies: a new "old" paradigm for HIV vaccines.* Clin Vaccine Immunol, 2014. **21**(8): p. 1023-36.

27. Tomaras, G.D., et al., *Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia.* J Virol, 2008. **82**(24): p. 12449-63.

28. Overbaugh, J. and L. Morris, *The Antibody Response against HIV-1.* Cold Spring Harb Perspect Med, 2012. **2**(1): p. a007039.

29. Walsh, S.R. and M.S. Seaman, *Broadly Neutralizing Antibodies for HIV-1 Prevention.* Front Immunol, 2021. **12**: p. 712122.

30. Julg, B., et al., *Protective Efficacy of Broadly Neutralizing Antibodies with Incomplete Neutralization Activity against Simian-Human Immunodeficiency Virus in Rhesus Monkeys.* J Virol, 2017. **91**(20).

31. Julg, B., et al., *Broadly neutralizing antibodies targeting the HIV-1 envelope V2 apex confer protection against a clade C SHIV challenge.* Sci Transl Med, 2017. **9**(406).

32. Moldt, B., et al., *Highly potent HIV-specific antibody neutralization in vitro translates into effective protection against mucosal SHIV challenge in vivo.* Proc Natl Acad Sci U S A, 2012. **109**(46): p. 18921-5.

33. Garber, D.A., et al., *Durable protection against repeated penile exposures to simian-human immunodeficiency virus by broadly neutralizing antibodies.* Nat Commun, 2020. **11**(1): p. 3195.

34. Gautam, R., et al., *A single injection of anti-HIV-1 antibodies protects against repeated SHIV challenges.* Nature, 2016. **533**(7601): p. 105-109.

35. Corey, L., et al., *Two Randomized Trials of Neutralizing Antibodies to Prevent HIV-1 Acquisition.* N Engl J Med, 2021. **384**(11): p. 1003-1014.

36. Haynes, B.F., et al., *Immune-correlates analysis of an HIV-1 vaccine efficacy trial.* N Engl J Med, 2012. **366**(14): p. 1275-86.

37. Hessell, A.J., et al., *Fc receptor but not complement binding is important in antibody protection against HIV.* Nature, 2007. **449**(7158): p. 101-4.

38. Milligan, C., et al., *Passively acquired antibody-dependent cellular cytotoxicity (ADCC) activity in HIV-infected infants is associated with reduced mortality.* Cell Host Microbe, 2015. **17**(4): p. 500-6.

39. Yaffe, Z.A., et al., *Improved HIV-positive infant survival is correlated with high levels of HIV-specific ADCC activity in multiple cohorts.* Cell Rep Med, 2021. **2**(4): p. 100254.

40. Rijkers, G., et al., *Differences in Antibody Kinetics and Functionality Between Severe and Mild Severe Acute Respiratory Syndrome Coronavirus 2 Infections.* J Infect Dis, 2020. **222**(8): p. 1265-1269.

41. Lau, E.H.Y., et al., *Neutralizing antibody titres in SARS-CoV-2 infections.* Nat Commun, 2021. **12**(1): p. 63.

42. Turner, J.S., et al., *SARS-CoV-2 infection induces long-lived bone marrow plasma cells in humans.* Nature, 2021. **595**(7867): p. 421-425.

43. McMahan, K., et al., *Correlates of protection against SARS-CoV-2 in rhesus macaques.* Nature, 2021. **590**(7847): p. 630-634.

44. Ullah, I., et al., *Live imaging of SARS-CoV-2 infection in mice reveals neutralizing antibodies require Fc function for optimal efficacy.* bioRxiv, 2021.

45. Schäfer, A., et al., *Antibody potency, effector function, and combinations in protection and therapy for SARS-CoV-2 infection in vivo.* J Exp Med, 2021. **218**(3).

46. Addetia, A., et al., *Neutralizing Antibodies Correlate with Protection from SARS-CoV-2 in Humans during a Fishery Vessel Outbreak with a High Attack Rate.* J Clin Microbiol, 2020. **58**(11).

47. Hanrath, A.T., B.A.I. Payne, and C.J.A. Duncan, *Prior SARS-CoV-2 infection is associated with protection against symptomatic reinfection.* J Infect, 2021. **82**(4): p. e29-e30.

48. Lumley, S.F., et al., *Antibody Status and Incidence of SARS-CoV-2 Infection in Health Care Workers.* N Engl J Med, 2021. **384**(6): p. 533-540.

49. Khoury, D.S., et al., *Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection.* Nat Med, 2021. **27**(7): p. 1205-1211.

50. Earle, K.A., et al., *Evidence for antibody as a protective correlate for COVID-19 vaccines.* Vaccine, 2021. **39**(32): p. 4423-4428.

51. Tauzin, A., et al., *A single dose of the SARS-CoV-2 vaccine BNT162b2 elicits Fc-mediated antibody effector functions and T cell responses.* Cell Host Microbe, 2021.

52. Eguia, R.T., et al., *A human coronavirus evolves antigenically to escape antibody immunity.* PLoS Pathog, 2021. **17**(4): p. e1009453.

53. Edridge, A.W.D., et al., *Seasonal coronavirus protective immunity is short-lasting.* Nat Med, 2020. **26**(11): p. 1691-1693.

54. Graham, B.S., M.S.A. Gilman, and J.S. McLellan, *Structure-Based Vaccine Antigen Design.* Annu Rev Med, 2019. **70**: p. 91-104.

55. Abbott, W.M., M.M. Damschroder, and D.C. Lowe, *Current approaches to fine mapping of antigen-antibody interactions.* Immunology, 2014. **142**(4): p. 526-35.

56. Fowler, D.M. and S. Fields, *Deep mutational scanning: a new style of protein science.* Nat Methods, 2014. **11**(8): p. 801-7.

57. Dingens, A.S., et al., *An Antigenic Atlas of HIV-1 Escape from Broadly Neutralizing Antibodies Distinguishes Functional and Structural Epitopes.* Immunity, 2019. **50**(2): p. 520-532.e3.

58. Dingens, A.S., et al., *Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody.* Cell Host Microbe, 2017. **21**(6): p. 777-787.e4.

59. Doud, M.B., S.E. Hensley, and J.D. Bloom, *Complete mapping of viral escape from neutralizing antibodies.* PLoS Pathog, 2017. **13**(3): p. e1006271.

60. Doud, M.B., J.M. Lee, and J.D. Bloom, *How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin.* Nat Commun, 2018. **9**(1): p. 1386.

61. Greaney, A.J., et al., *Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies.* Cell Host Microbe, 2021.

62. Greaney, A.J., et al., *Antibodies elicited by mRNA-1273 vaccination bind more broadly to the receptor binding domain than do those from SARS-CoV-2 infection.* Sci Transl Med, 2021. **13**(600).

63. Mohan, D., et al., *PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes.* Nat Protoc, 2018. **13**(9): p. 1958-1978.

64. Williams, K.L., et al., *Identification of HIV gp41-specific antibodies that mediate killing of infected cells.* PLoS Pathog, 2019. **15**(2): p. e1007572.

65. Doepker, L.E., et al., *Diversity and function of maternal HIV-1-specific antibodies at the time of vertical transmission.* J Virol, 2020.

66. Finton, K.A., et al., *Ontogeny of recognition specificity and functionality for the broadly neutralizing anti-HIV antibody 4E10.* PLoS Pathog, 2014. **10**(9): p. e1004403.

67. Finton, K.A., et al., *Autoreactivity and exceptional CDR plasticity (but not unusual polyspecificity) hinder elicitation of the anti-HIV antibody 4E10.* PLoS Pathog, 2013. **9**(9): p. e1003639.

68. Xu, G.J., et al., *Comprehensive serological profiling of human populations using a synthetic human virome.* Science, 2015. **348**(6239): p. aaa0698.

69. Larman, H.B., et al., *Autoantigen discovery with a synthetic human peptidome.* Nat Biotechnol, 2011. **29**(6): p. 535-41.

70. Fowler, D.M., et al., *High-resolution mapping of protein sequence-function relationships.* Nat Methods, 2010. **7**(9): p. 741-6.

71. Starita, L.M., et al., *Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis.* Proc Natl Acad Sci U S A, 2013. **110**(14): p. E1263-72.

72. Ernst, A., et al., *Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing.* Mol Biosyst, 2010. **6**(10): p. 1782-90.

73. Zinkus-Boltz, J., C. DeValk, and B.C. Dickinson, *A Phage-Assisted Continuous Selection Approach for Deep Mutational Scanning of Protein-Protein Interactions.* ACS Chem Biol, 2019. **14**(12): p. 2757-2767.

74. Hemelaar, J., et al., *Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis.* Lancet Infect Dis, 2019. **19**(2): p. 143-155.

75. Xu, J.Y., et al., *Epitope mapping of two immunodominant domains of gp41, the transmembrane protein of human immunodeficiency virus type 1, using ten human monoclonal antibodies.* J Virol, 1991. **65**(9): p. 4832-8.

76. Cavacini, L.A., et al., *Functional and molecular characterization of human monoclonal antibody reactive with the immunodominant region of HIV type 1 glycoprotein 41.* AIDS Res Hum Retroviruses, 1998. **14**(14): p. 1271-80.

77. Gorny, M.K., et al., *Neutralization of diverse human immunodeficiency virus type 1 variants by an anti-V3 human monoclonal antibody.* J Virol, 1992. **66**(12): p. 7538-42.

78. Gorny, M.K., et al., *Production of site-selected neutralizing human monoclonal antibodies against the third variable domain of the human immunodeficiency virus type 1 envelope glycoprotein.* Proc Natl Acad Sci U S A, 1991. **88**(8): p. 3238-42.

79. Gohain, N., et al., *Molecular basis for epitope recognition by non-neutralizing anti-gp41 antibody F240.* Sci Rep, 2016. **6**: p. 36685.

80. Gorny, M.K., et al., *Repertoire of neutralizing human monoclonal antibodies specific for the V3 domain of HIV-1 gp120.* J Immunol, 1993. **150**(2): p. 635-43.

81. Burke, V., et al., *Structural basis of the cross-reactivity of genetically related human anti-HIV-1 mAbs: implications for design of V3-based immunogens.* Structure, 2009. **17**(11): p. 1538-46.

82. Patel, M.B., N.G. Hoffman, and R. Swanstrom, *Subtype-specific conformational differences within the V3 region of subtype B and subtype C human immunodeficiency virus type 1 Env proteins.* J Virol, 2008. **82**(2): p. 903-16.

83. Guido, N., et al., *Improved PCR Amplification of Broad Spectrum GC DNA Templates.* PLoS One, 2016. **11**(6): p. e0156478.

84. Stanfield, R.L., et al., *Structural rationale for the broad neutralization of HIV-1 by human monoclonal antibody 447-52D.* Structure, 2004. **12**(2): p. 193-204.

85. Falkowska, E., et al., *PGV04, an HIV-1 gp120 CD4 binding site antibody, is broad and potent in neutralization but does not induce conformational changes characteristic of CD4.* J Virol, 2012. **86**(8): p. 4394-403.

86. Kelley, R.F. and M.P. O'Connell, *Thermodynamic analysis of an antibody functional epitope.* Biochemistry, 1993. **32**(27): p. 6828-35.

87. Li, Y., et al., *Mechanism of neutralization by the broadly neutralizing HIV-1 monoclonal antibody VRC01.* J Virol, 2011. **85**(17): p. 8954-67.

88. Jin, L., B.M. Fendly, and J.A. Wells, *High resolution functional analysis of antibody-antigen interactions.* J Mol Biol, 1992. **226**(3): p. 851-65.

89. Muller, Y.A., et al., *VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 A resolution and mutational analysis of the interface.* Structure, 1998. **6**(9): p. 1153-67.

90. Dall'Acqua, W., et al., *A mutational analysis of binding interactions in an antigen-antibody protein-protein complex.* Biochemistry, 1998. **37**(22): p. 7981-91.

91. Pons, J., A. Rajpal, and J.F. Kirsch, *Energetic analysis of an antigen/antibody interface: alanine scanning mutagenesis and double mutant cycles on the HyHEL-10/lysozyme interaction.* Protein Sci, 1999. **8**(5): p. 958-68.

92. Kowalsky, C.A., et al., *Rapid fine conformational epitope mapping using comprehensive mutagenesis and deep sequencing.* J Biol Chem, 2015. **290**(44): p. 26457-70.

93. Van Blarcom, T., et al., *Precise and efficient antibody epitope determination through library design, yeast display and next-generation sequencing.* J Mol Biol, 2015. **427**(6 Pt B): p. 1513-1534.

94. Sheehan, J. and W.A. Marasco, *Phage and Yeast Display.* Microbiol Spectr, 2015. **3**(1): p. Aid-0028-2014.

95. Chen, X., J.L. Zaro, and W.C. Shen, *Fusion protein linkers: property, design and functionality.* Adv Drug Deliv Rev, 2013. **65**(10): p. 1357-69.

96. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.

97. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.* Cell, 2020. **181**(2): p. 281-292.e6.

98. Xia, S., et al., *Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion.* Cell Res, 2020. **30**(4): p. 343-355.

99. Walls, A.C., et al., *Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion.* Proc Natl Acad Sci U S A, 2017. **114**(42): p. 11157-11162.

100. Belouzard, S., V.C. Chu, and G.R. Whittaker, *Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites.* Proc Natl Acad Sci U S A, 2009. **106**(14): p. 5871-6.

101. Fan, X., et al., *Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein.* Nat Commun, 2020. **11**(1): p. 3618.

102. Wang, C., et al., *A human monoclonal antibody blocking SARS-CoV-2 infection.* Nat Commun, 2020. **11**(1): p. 2251.

103. Pinto, D., et al., *Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody.* Nature, 2020. **583**(7815): p. 290-295.

104. Wec, A.Z., et al., *Broad neutralization of SARS-related viruses by human monoclonal antibodies.* Science, 2020. **369**(6504): p. 731-736.

105. Wan, J., et al., *Human-IgG-Neutralizing Monoclonal Antibodies Block the SARS-CoV-2 Infection.* Cell Rep, 2020. **32**(3): p. 107918.

106. Rogers, T.F., et al., *Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model.* Science, 2020. **369**(6506): p. 956-963.

107. Hassan, A.O., et al., *A SARS-CoV-2 Infection Model in Mice Demonstrates Protection by Neutralizing Antibodies.* Cell, 2020. **182**(3): p. 744-753.e4.

108. Zost, S.J., et al., *Potently neutralizing and protective human antibodies against SARS-CoV-2.* Nature, 2020. **584**(7821): p. 443-449.

109. Shanmugaraj, B., et al., *Perspectives on monoclonal antibody therapy as potential therapeutic intervention for Coronavirus disease-19 (COVID-19).* Asian Pac J Allergy Immunol, 2020. **38**(1): p. 10-18.

110. Poh, C.M., et al., *Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients.* Nat Commun, 2020. **11**(1): p. 2806.

111. Li, Y., et al., *Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients.* Cell Mol Immunol, 2020.

112. Saphire, E.O., et al., *Antibody-mediated protection against Ebola virus.* Nat Immunol, 2018. **19**(11): p. 1169-1178.

113. Gunn, B.M., et al., *A Role for Fc Function in Therapeutic Monoclonal Antibody-Mediated Protection against Ebola Virus.* Cell Host Microbe, 2018. **24**(2): p. 221-233.e5.

114. Robbiani, D.F., et al., *Convergent antibody responses to SARS-CoV-2 in convalescent individuals.* Nature, 2020. **584**(7821): p. 437-442.

115. Sui, J., et al., *Broadening of neutralization activity to directly block a dominant antibody-driven SARS-coronavirus evolution pathway.* PLoS Pathog, 2008. **4**(11): p. e1000197.

116. Kleine-Weber, H., et al., *Mutations in the Spike Protein of Middle East Respiratory Syndrome Coronavirus Transmitted in Korea Increase Resistance to Antibody-Mediated Neutralization.* J Virol, 2019. **93**(2).

117. Li, Q., et al., *The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity.* Cell, 2020. **182**(5): p. 1284-1294.e9.

118. Starr, T.N., et al., *Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding.* Cell, 2020. **182**(5): p. 1295-1310.e20.

119. Garrett, M.E., et al., *Phage-DMS: A Comprehensive Method for Fine Mapping of Antibody Epitopes.* iScience, 2020. **23**(10).

120. Korber, B., et al., *Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus.* Cell, 2020. **182**(4): p. 812-827.e19.

121. Hilton, S.K., et al., *dms-view: Interactive visualization tool for deep mutational scanning data.* Journal of Open Source Software, 2020. **5**(52).

122. Crawford, K.H.D., et al., *Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays.* Viruses, 2020. **12**(5).

123. Piccoli, L., et al., *Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology.* Cell, 2020. **183**(4): p. 1024-1042.e21.

124. Jiang, H.W., et al., *SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses.* Nat Commun, 2020. **11**(1): p. 3581.

125. Shrock, E., et al., *Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity.* Science, 2020.

126. Zamecnik, C.R., et al., *ReScan, a Multiplex Diagnostic Pipeline, Pans Human Sera for SARS-CoV-2 Antigens.* Cell Rep Med, 2020: p. 100123.

127. Yi, Z., et al., *Functional mapping of B-cell linear epitopes of SARS-CoV-2 in COVID-19 convalescent population.* Emerg Microbes Infect, 2020. **9**(1): p. 1988-1996.

128. Seydoux, E., et al., *Analysis of a SARS-CoV-2-Infected Individual Reveals Development of Potent Neutralizing Antibodies with Limited Somatic Mutation.* Immunity, 2020. **53**(1): p. 98-105.e5.

129. Brouwer, P.J.M., et al., *Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability.* Science, 2020. **369**(6504): p. 643-650.

130. Patel, J.S., et al., *Expanding the watch list for potential Ebola virus antibody escape mutations.* PLoS One, 2019. **14**(3): p. e0211093.

131. Domingo, E. and J.J. Holland, *RNA virus mutations and fitness for survival.* Annu Rev Microbiol, 1997. **51**: p. 151-78.

132. Louie, R.H.Y., et al., *Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies.* Proc Natl Acad Sci U S A, 2018. **115**(4): p. E564-e573.

133. Kosik, I., et al., *Influenza A virus hemagglutinin glycosylation compensates for antibody escape fitness costs.* PLoS Pathog, 2018. **14**(1): p. e1006796.

134. Keck, Z.Y., et al., *Mutations in hepatitis C virus E2 located outside the CD81 binding sites lead to escape from broadly neutralizing antibodies but compromise virus infectivity.* J Virol, 2009. **83**(12): p. 6149-60.

135. Lip, K.M., et al., *Monoclonal antibodies targeting the HR2 domain and the region immediately upstream of the HR2 of the S protein neutralize in vitro infection of severe acute respiratory syndrome coronavirus.* J Virol, 2006. **80**(2): p. 941-50.

136. Zheng, Z., et al., *Monoclonal antibodies for the S2 subunit of spike of SARS-CoV-1 cross-react with the newly-emerged SARS-CoV-2.* Euro Surveill, 2020. **25**(28).

137. Baum, A., et al., *Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies.* Science, 2020. **369**(6506): p. 1014-1018.

138. Mabuka, J., et al., *HIV-specific antibodies capable of ADCC are common in breastmilk and are associated with reduced risk of transmission in women with high viral loads.* PLoS Pathog, 2012. **8**(6): p. e1002739.

139. Dingens, A.S., et al., *Serological identification of SARS-CoV-2 infections among children visiting a hospital during the initial Seattle outbreak.* Nat Commun, 2020. **11**(1): p. 4378.

140. Stadlbauer, D., et al., *SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a Serological Assay, Antigen Production, and Test Setup.* Curr Protoc Microbiol, 2020. **57**(1): p. e100.

141. Polack, F.P., et al., *Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine.* N Engl J Med, 2020. **383**(27): p. 2603-2615.

142. Keehner, J., et al., *SARS-CoV-2 Infection after Vaccination in Health Care Workers in California.* N Engl J Med, 2021. **384**(18): p. 1774-1775.

143. Amit, S., et al., *Early rate reductions of SARS-CoV-2 infection and COVID-19 in BNT162b2 vaccine recipients.* Lancet, 2021. **397**(10277): p. 875-877.

144. Angel, Y., et al., *Association Between Vaccination With BNT162b2 and Incidence of Symptomatic and Asymptomatic SARS-CoV-2 Infections Among Health Care Workers.* Jama, 2021.

145. Thompson, M.G., et al., *Interim Estimates of Vaccine Effectiveness of BNT162b2 and mRNA-1273 COVID-19 Vaccines in Preventing SARS-CoV-2 Infection Among Health Care Personnel, First Responders, and Other Essential and Frontline Workers - Eight U.S. Locations, December 2020-March 2021.* MMWR Morb Mortal Wkly Rep, 2021. **70**(13): p. 495-500.

146. Haas, E.J., et al., *Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data.* Lancet, 2021.

147. Baden, L.R., et al., *Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine.* N Engl J Med, 2021. **384**(5): p. 403-416.

148. Planas, D., et al., *Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization.* Nature, 2021.

149. Lopez Bernal, J., et al., *Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant.* N Engl J Med, 2021.

150. Hendley, J.O., H.B. Fishburne, and J.M. Gwaltney, Jr., *Coronavirus infections in working adults. Eight-year study with 229 E and OC 43.* Am Rev Respir Dis, 1972. **105**(5): p. 805-11.

151. Schmidt, O.W., et al., *Rises in titers of antibody to human coronaviruses OC43 and 229E in Seattle families during 1975-1979.* Am J Epidemiol, 1986. **123**(5): p. 862-8.

152. Li, Y., et al., *Linear epitope landscape of the SARS-CoV-2 Spike protein constructed from 1,051 COVID-19 patients.* Cell Rep, 2021. **34**(13): p. 108915.

153. Stoddard, C.I., et al., *Epitope profiling reveals binding signatures of SARS-CoV-2 immune response in natural infection and cross-reactivity with endemic human CoVs.* Cell Rep, 2021. **35**(8): p. 109164.

154. Goel, R.R., et al., *Distinct antibody and memory B cell responses in SARS-CoV-2 naïve and recovered individuals following mRNA vaccination.* Sci Immunol, 2021. **6**(58).

155. Prendecki, M., et al., *Effect of previous SARS-CoV-2 infection on humoral and T-cell responses to single-dose BNT162b2 vaccine.* Lancet, 2021. **397**(10280): p. 1178-1181.

156. Edara, V.V., et al., *Neutralizing Antibodies Against SARS-CoV-2 Variants After Infection and Vaccination.* Jama, 2021.

157. Garrett, M.E., et al., *High-resolution profiling of pathways of escape for SARS-CoV-2 spike-binding antibodies.* Cell, 2021. **184**(11): p. 2927-2938.e11.

158. Winkler, E.S., et al., *Human neutralizing antibodies against SARS-CoV-2 require intact Fc effector functions for optimal therapeutic protection.* Cell, 2021. **184**(7): p. 1804-1820.e16.

159. Jackson, L.A., et al., *An mRNA Vaccine against SARS-CoV-2 - Preliminary Report.* N Engl J Med, 2020. **383**(20): p. 1920-1931.

160. Sauer, M.M., et al., *Structural basis for broad coronavirus neutralization.* Nat Struct Mol Biol, 2021. **28**(6): p. 478-486.

161. Gaebler, C., et al., *Evolution of antibody immunity to SARS-CoV-2.* Nature, 2021. **591**(7851): p. 639-644.

162. Sakharkar, M., et al., *Prolonged evolution of the human B cell response to SARS-CoV-2 infection.* Sci Immunol, 2021. **6**(56).

163. Turner, J.S., et al., *SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses.* Nature, 2021.

164. Greaney, A.J., F.C. Welsh, and J.D. Bloom, *Co-dominant neutralizing epitopes make anti-measles immunity resistant to viral evolution.* Cell Rep Med, 2021. **2**(4): p. 100257.

165. Jaroszewski, L., et al., *The interplay of SARS-CoV-2 evolution and constraints imposed by the structure and functionality of its proteins.* PLoS Comput Biol, 2021. **17**(7): p. e1009147.

166. Di Tommaso, P., et al., *Nextflow enables reproducible computational workflows.* Nat Biotechnol, 2017. **35**(4): p. 316-319.

167. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.

168. Pedregosa, F., et al., *Scikit-learn: Machine Learning in Python.* J. Mach. Learn. Res., 2011. **12**(null): p. 2825–2830.

169. Pinto, D., et al., *Broad betacoronavirus neutralization by a stem helix-specific human antibody.* Science, 2021.

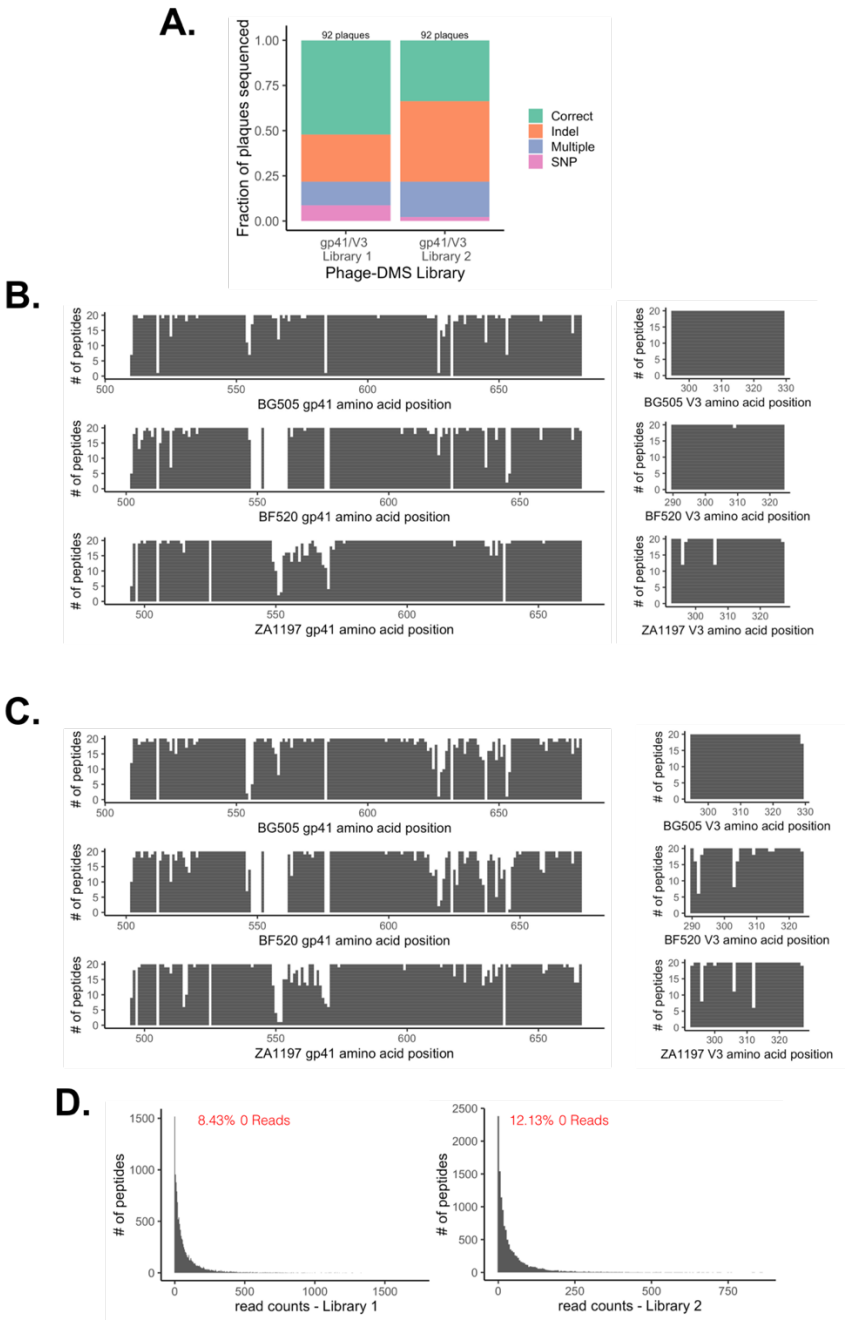
170. Chen, E.C., et al., *Convergent antibody responses to the SARS-CoV-2 spike protein in convalescent and vaccinated individuals.* Cell Reports, 2021.

171. Xu, K., et al., *Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1.* Nat Med, 2018. **24**(6): p. 857-867.

Appendix A

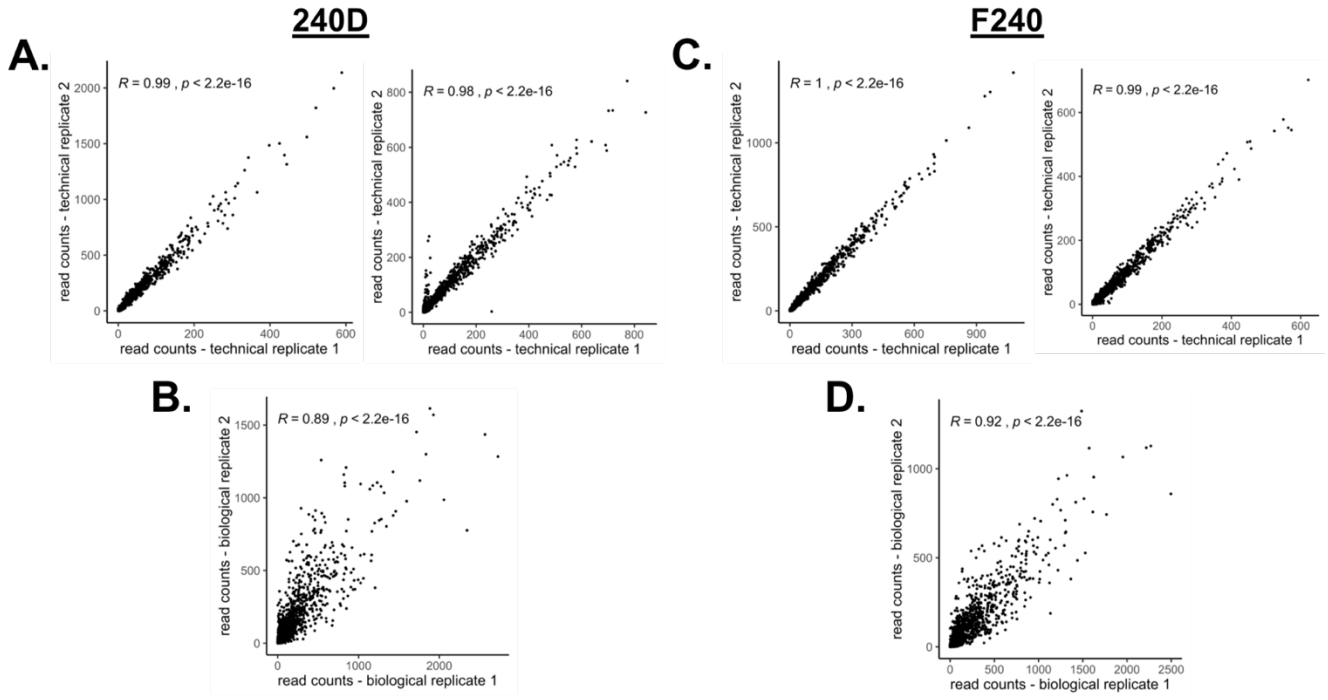
SUPPLEMENTARY FILES FOR CHAPTER 3

SUPPLEMENTAL FIGURES

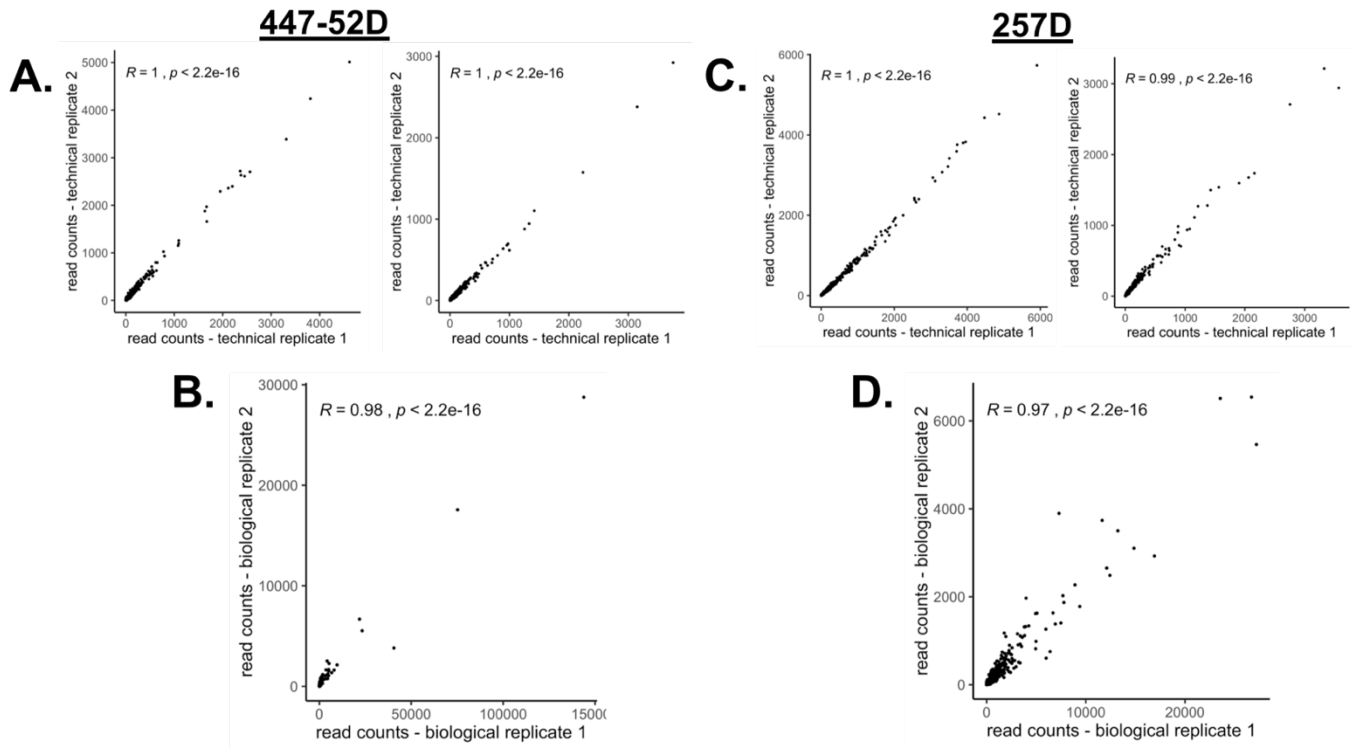


Supplemental Figure 2.1. Composition and distribution of sequences in the gp41/V3 Phage-DMS libraries. Related to Figure 2.2 and Figure 2.3. (A) Stacked bar plot depicting

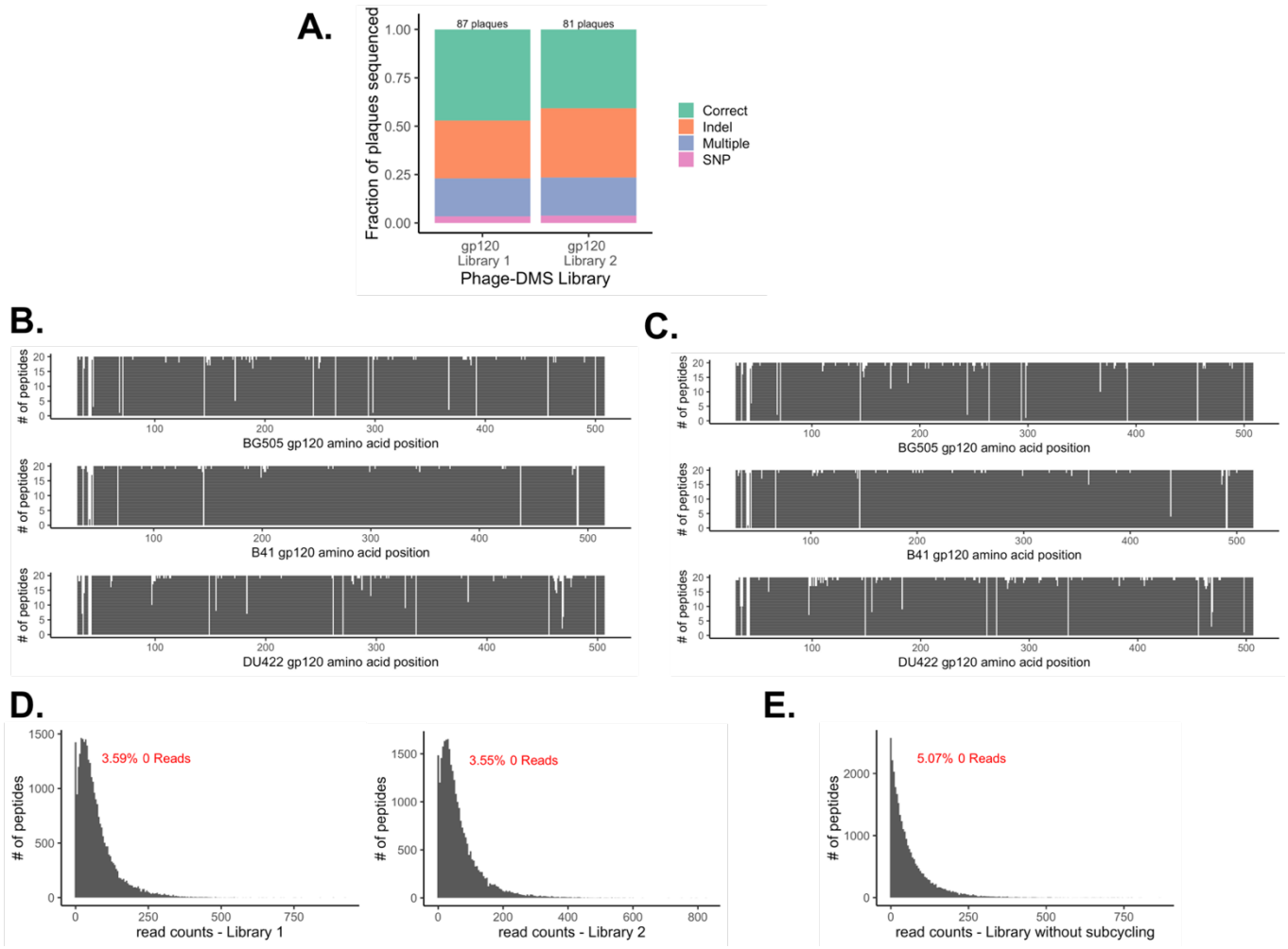
results of Sanger sequencing individual plaques from each library. Total number of plaques sequenced is displayed at the top. SNP = single nucleotide polymorphism. **(B)** Coverage of all unique sequences centered at sites across gp41 and V3 within gp41/V3 library 1, as determined by deep sequencing at 82-fold coverage. The height of the line at each site corresponds to whether each unique amino acid variant was counted at least once, with 20 total possible amino acids. **(C)** Coverage of all unique sequences centered at sites across gp41 and V3 within gp41/V3 library 2, as determined by deep sequencing at 45-fold coverage. **(D)** Histogram showing the distribution of all reads sequenced from gp41/V3 library 1 and 2 as determined by deep sequencing. Percent of computationally designed sequences not counted shown in red.



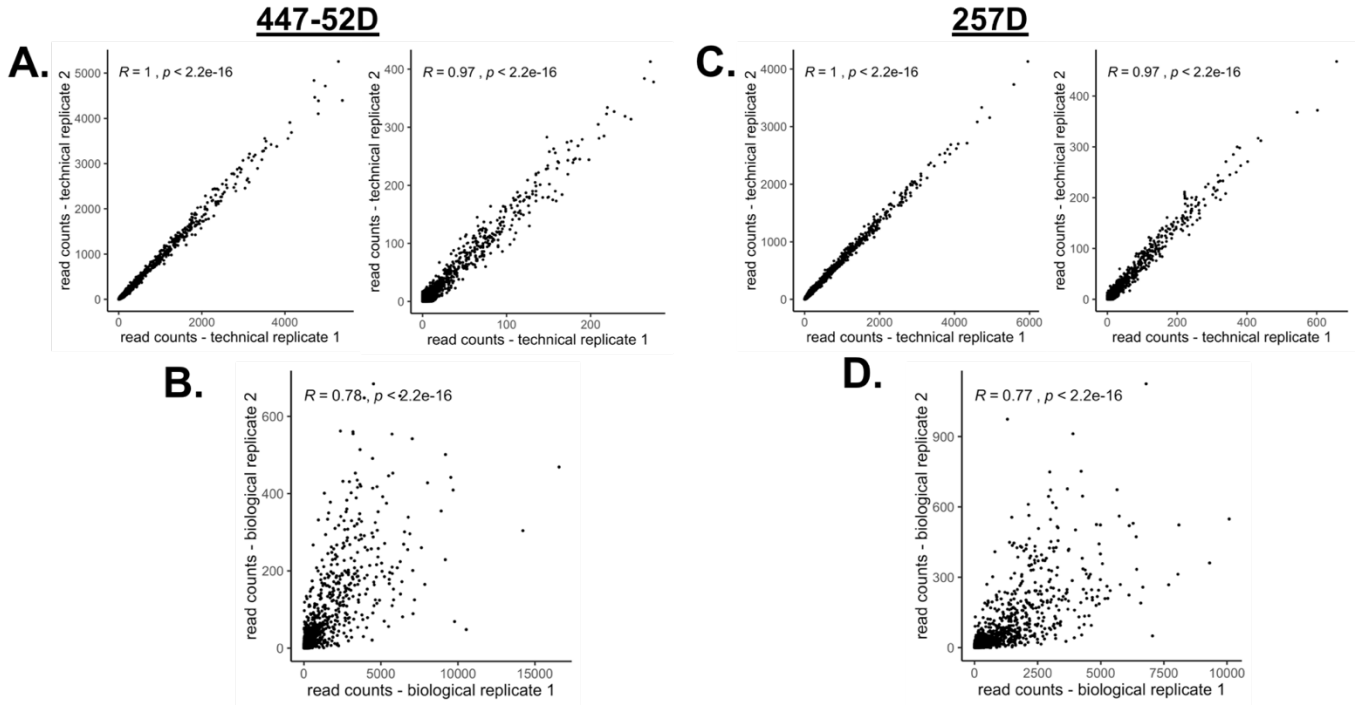
Supplemental Figure 2.2. Reproducibility of gp41/V3 Phage-DMS experiments with gp41-specific antibodies. Related to Figure 2.2. Correlation of raw read counts between technical and biological replicates for mAb 240D and mAb F240. **(A, C)** Correlation between replicate wells done in parallel for each experiment. **(B, D)** Correlation between separate experiments performed with gp41/V3 library 1 and 2. Pearson's correlation is shown at the top.



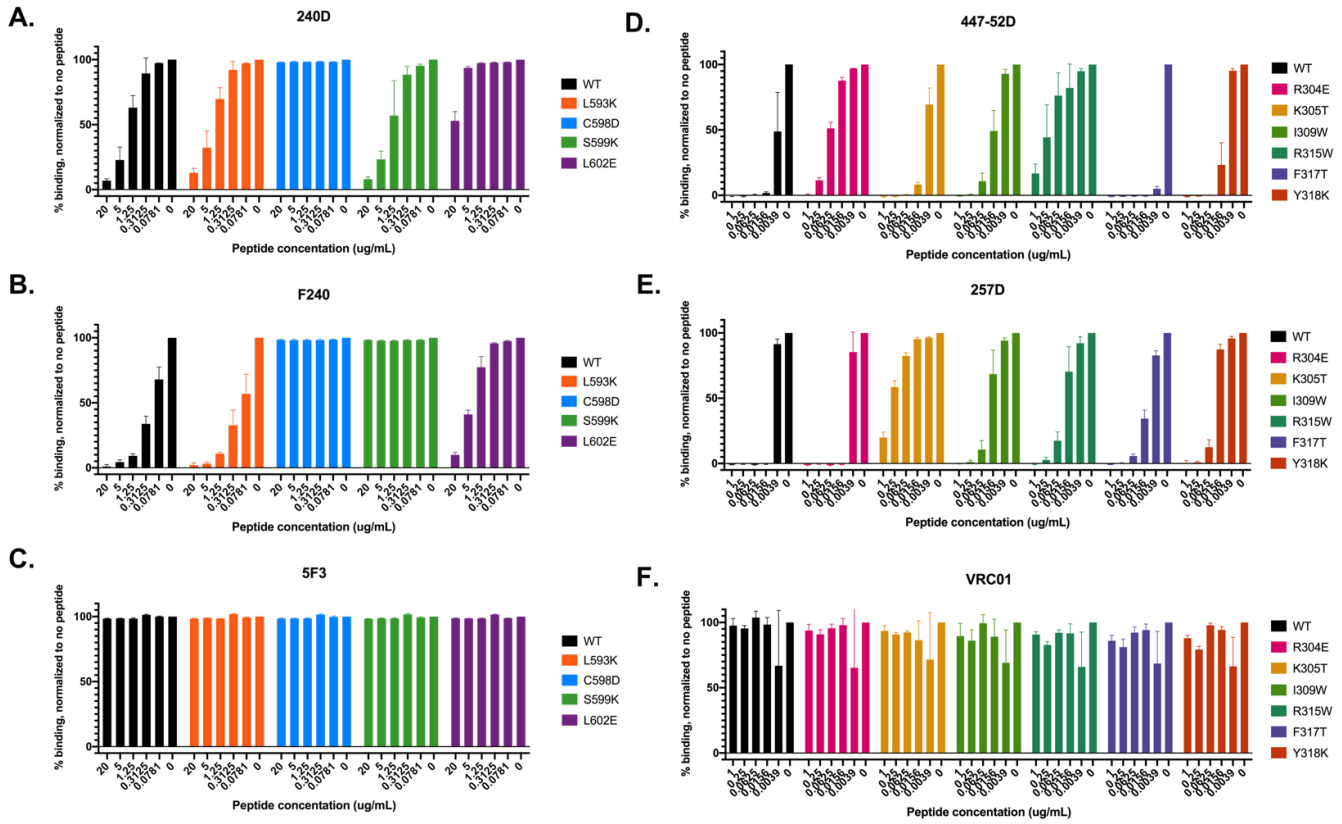
Supplemental Figure 2.3. Reproducibility of gp41/V3 Phage-DMS experiments with V3-specific antibodies. Related to Figure 2.3. Correlation of raw read counts between technical and biological replicates for mAb 447-52D and mAb 257D. **(A, C)** Correlation between replicate wells done in parallel for each experiment. **(B, D)** Correlation between separate experiments performed with gp41/V3 library 1 and 2. Pearson's correlation is shown at the top.



Supplemental Figure 2.4. Composition and distribution of sequences in the gp120 Phage-DMS libraries. **Related to Figure 2.4. (A)** Stacked bar plot depicting results of Sanger sequencing individual plaques from each library. Total number of plaques sequenced is displayed at the top. SNP = single nucleotide polymorphism. **(B)** Coverage of all unique sequences centered at sites across gp120 within gp120 library 1, as determined by deep sequencing at 71-fold coverage. The height of the line at each site corresponds to whether each unique amino acid variant was counted at least once, with 20 total possible. **(C)** Coverage of all unique sequences centered at sites across gp120 within gp120 library 2, as determined by deep sequencing at 65-fold coverage. **(D)** Histogram showing the distribution of all reads sequenced from gp120 library 1 and 2 as determined by deep sequencing. Percent of computationally designed sequences not counted shown in red. **(E)** Histogram showing the distribution of all reads sequenced from a gp120 library amplified without using subcycling PCR. Percent of computationally designed sequences not counted shown in red.



Supplemental Figure 2.5. Reproducibility of gp120 Phage-DMS experiments with V3-specific antibodies. Related to Figure 2.4. Correlation of raw read counts between technical and biological replicates for mAb 447-52D and mAb 257D. **(A, C)** Correlation between replicate wells done in parallel for each experiment. **(B, D)** Correlation between separate experiments performed with gp120 library 1 and 2. Pearson's correlation is shown at the top.



Supplemental Figure 2.6. Results of competition peptide ELISAs. Related to Figure 2.5. **(A-C)** Bar plots show ability of wild type and mutant peptides to block binding of (A) mAb 240D, (B) mAb F240, or (C) mAb 5F3 to MN gp41 protein. 5F3 is a control antibody known to bind to gp41 outside of the C-C loop region. Peptide concentrations are show below each group and represent 4-fold serial dilutions beginning at 20 ug/mL. Gp41-specific antibodies used at a concentration of 0.5 ug/mL, and gp41 protein was coated at 0.5 ug/mL. For reference, the gp41 sequence of MN corresponding to the region the gp41 peptides span is LLGFWGCSGKLICTTTVP. **(D-F)** Bar plots show ability of wild type and mutant peptides to block binding of (D) mAb 447-52D, (E) mAb 257D, or (F) mAb VRC01 to MN gp41 protein. VRC01 is a control antibody known to bind to gp120 outside of the V3 loop region. Peptide concentrations are show below each group and represent 4-fold serial dilutions beginning at 20 ug/mL.V3-specific antibodies used at a concentration of 1 ug/mL, and gp41 protein was coated at 1 ug/mL. For reference, the V3 sequence of SF162 corresponding to the region the V3 peptides span is RKSITIGPGRAFY.

SUPPLEMENTAL TABLES

Phage-DMS library	Common name	Strain	Codon optimized?	GC optimized?	Source	GenBank #	Env AA sites included in library (autologous numbering)
Gp41/V3	BG505	BG505.W6M.ENV.C2	yes	no	Xu et al	DQ208458.1	295 - 329, 509 - 681
Gp41/V3	BF520	BF520.W14M.C2	yes	no	Simonich et al	KX168094.1	290 - 324, 501 - 673
Gp41/V3	ZA1197	C.ZA.1197MB	yes	no	Rousseau et al	AY463234.1	293 - 327, 494 - 666
Gp120	BG505	BG505.W6M.ENV.C2	yes	yes	Xu et al	DQ208458.1	30 - 508
Gp120	B41	9032-08.A1.4685	yes	yes	Keele et al	EU576114.1	30 - 515
Gp120	DU422	Du422.1	yes	yes	Li et al	DQ411854.1	30 - 506

Supplemental Table 2.1. Description of strains use to construct Phage-DMS libraries. Related to Figure 2.1. Listed are the strains of HIV Env that were used to generate tiling wild type or mutant sequences for either the gp41/V3 or gp120 Phage-DMS libraries.

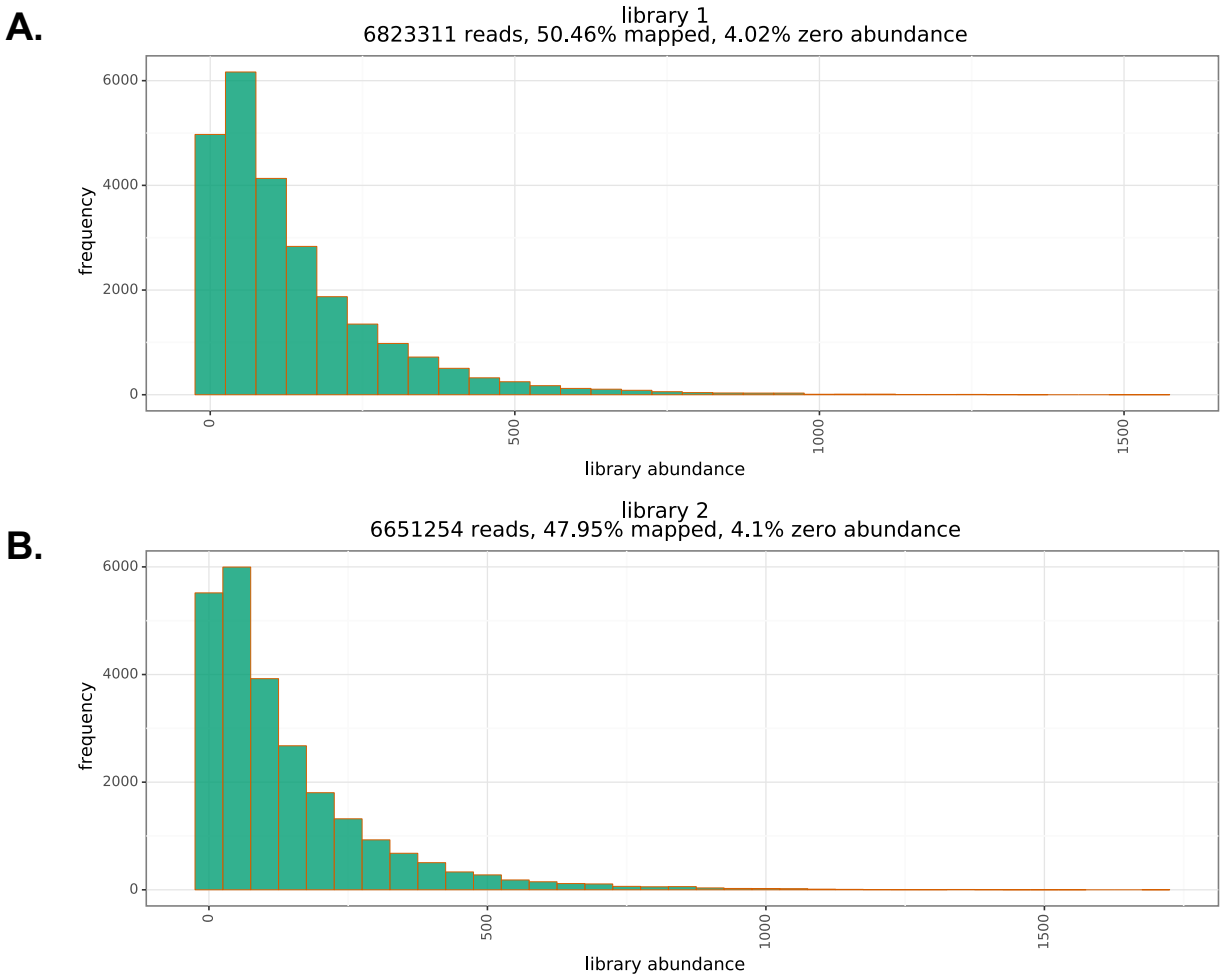
Peptide target	Strain	Env positions (autologous numbering)	Env positions (HXB2 numbering)	Mutation (HXB2 numbering)	Amino acid sequence
Gp41	BG505	589 – 606	592 – 609	WT	LLGIWGCSGKLICTTNVP
Gp41	BG505	589 – 606	592 – 609	L593K	L**K**GIWGCSGKLICTTNVP
Gp41	BG505	589 – 606	592 – 609	C598D	LLGIWG**D**SGKLICTTNVP
Gp41	BG505	589 – 606	592 – 609	S599K	LLGIWGC**K**GKLICTTNVP
Gp41	BG505	589 – 606	592 – 609	L602E	LLGIWGCSGK**E**ICTTNVP
V3	B41	313 – 325	304 – 318	WT	RKSIHIGPGRAFY
V3	B41	313 – 325	304 – 318	R304E	**E**KSIHIGPGRAFY
V3	B41	313 – 325	304 – 318	K305T	R**T**SIHIGPGRAFY
V3	B41	313 – 325	304 – 318	I309W	RKSIH**W**GPGRAFY
V3	B41	313 – 325	304 – 318	R315W	RKSIHIGPG**W**AFY
V3	B41	313 – 325	304 – 318	F317T	RKSIHIGPGRA**T**Y
V3	B41	313 – 325	304 – 318	Y318K	RKSIHIGPGRAF**K**

Supplemental Table 2.2. Description of peptides used to perform competition ELISAs. Related to Figure 2.5. Mutations are highlighted in red.

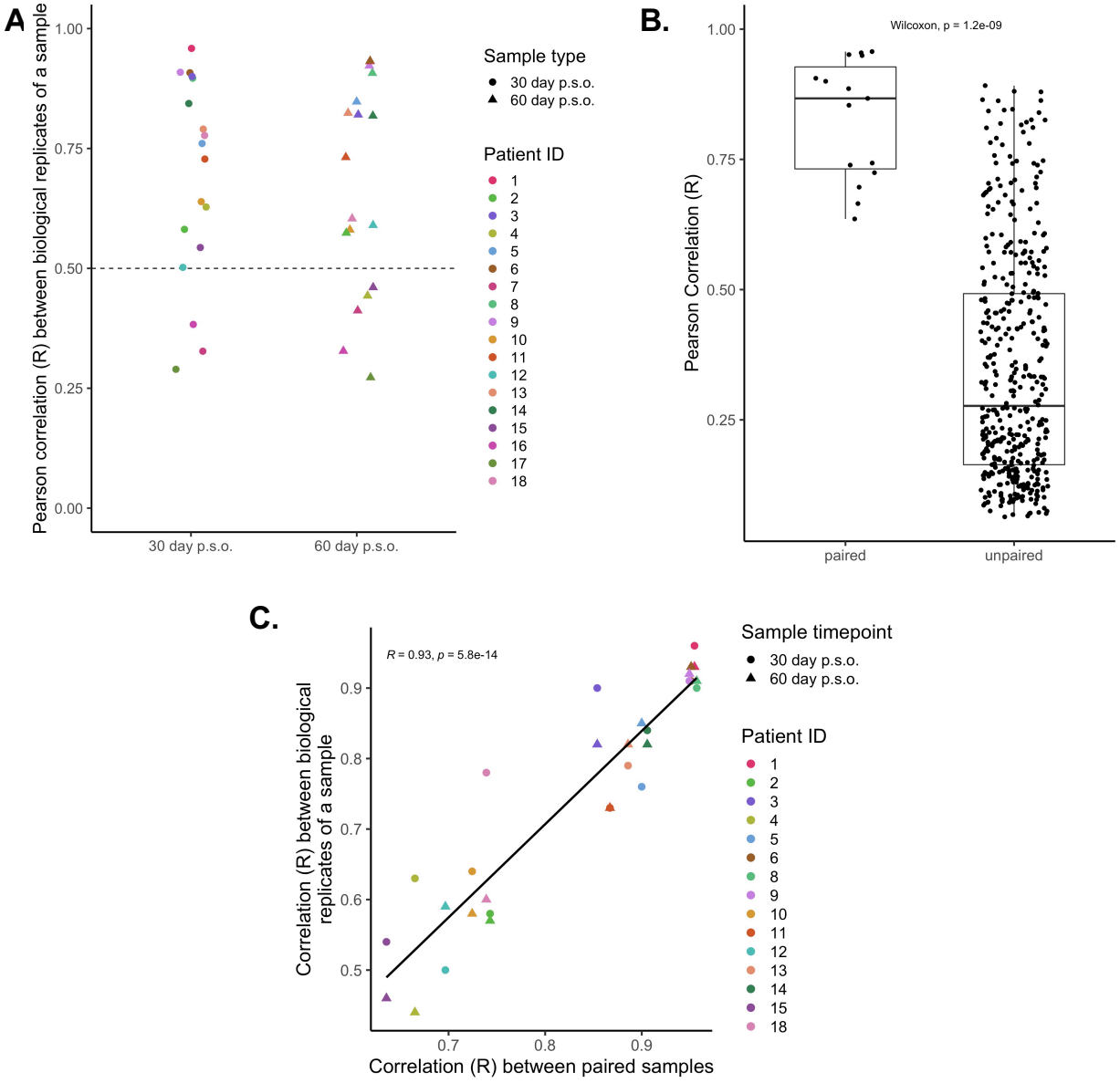
Appendix B

SUPPLEMENTARY FILES FOR CHAPTER 3

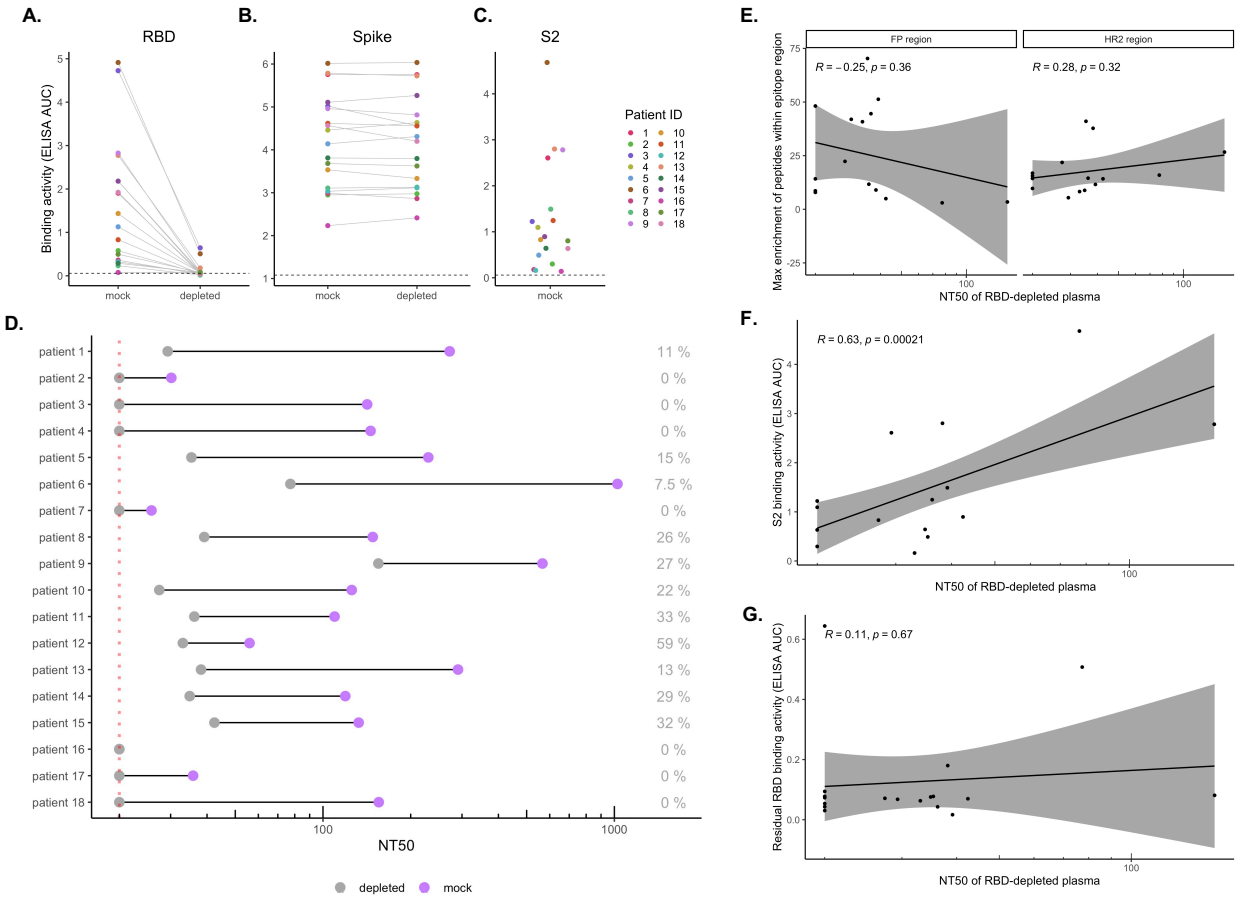
SUPPLEMENTAL FIGURES



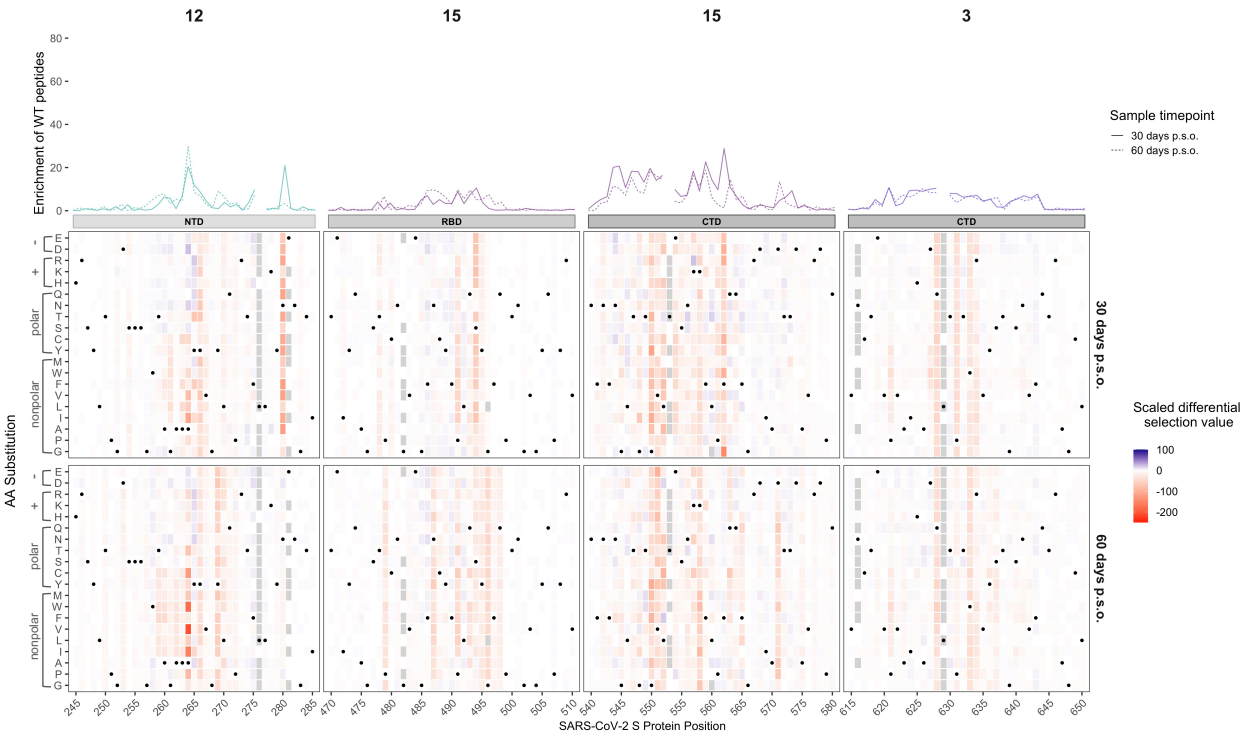
Supplemental Figure 3.1. Distribution of sequenced peptides within biological replicate Spike Phage-DMS libraries, related to Figure 3.1 and 3.2. (A and B) Histogram showing the distribution of all sequenced peptides from a representative deep sequencing experiment for Spike Phage-DMS Library 1 (A) and Library 2 (B). Reads were stringently aligned to the reference library, allowing for 0 mismatches, and the proportion of unmapped reads is shown at the top. Additionally, the proportion of all non-sequenced peptides for each library is shown at the top.



Supplemental Figure 3.2. Reproducibility of peptide enrichment by plasma from COVID-19 patients, related to Figure 3.2. (A) Distribution of correlation values between peptide enrichment values for replicate experiments with samples from COVID-19 patients (Pearson's correlation coefficient, R). Each color corresponds to a unique patient or volunteer, and the shape of each dot represents the type of sample. A dotted line at y = 0.5 represents the cutoff used to determine whether samples were kept in the analysis. (B) Boxplots showing the distribution of correlation values between patient samples that were paired between the day 30 and 60 p.s.o. timepoints (on the left) or samples that were randomly paired and compared (on the right). (C) Relationship between the biological replicate correlation for a sample and its correlation with its paired timepoint. Each color corresponds to a unique patient, and the shape of each dot represents the type of sample. Pearson's correlation coefficient shown.



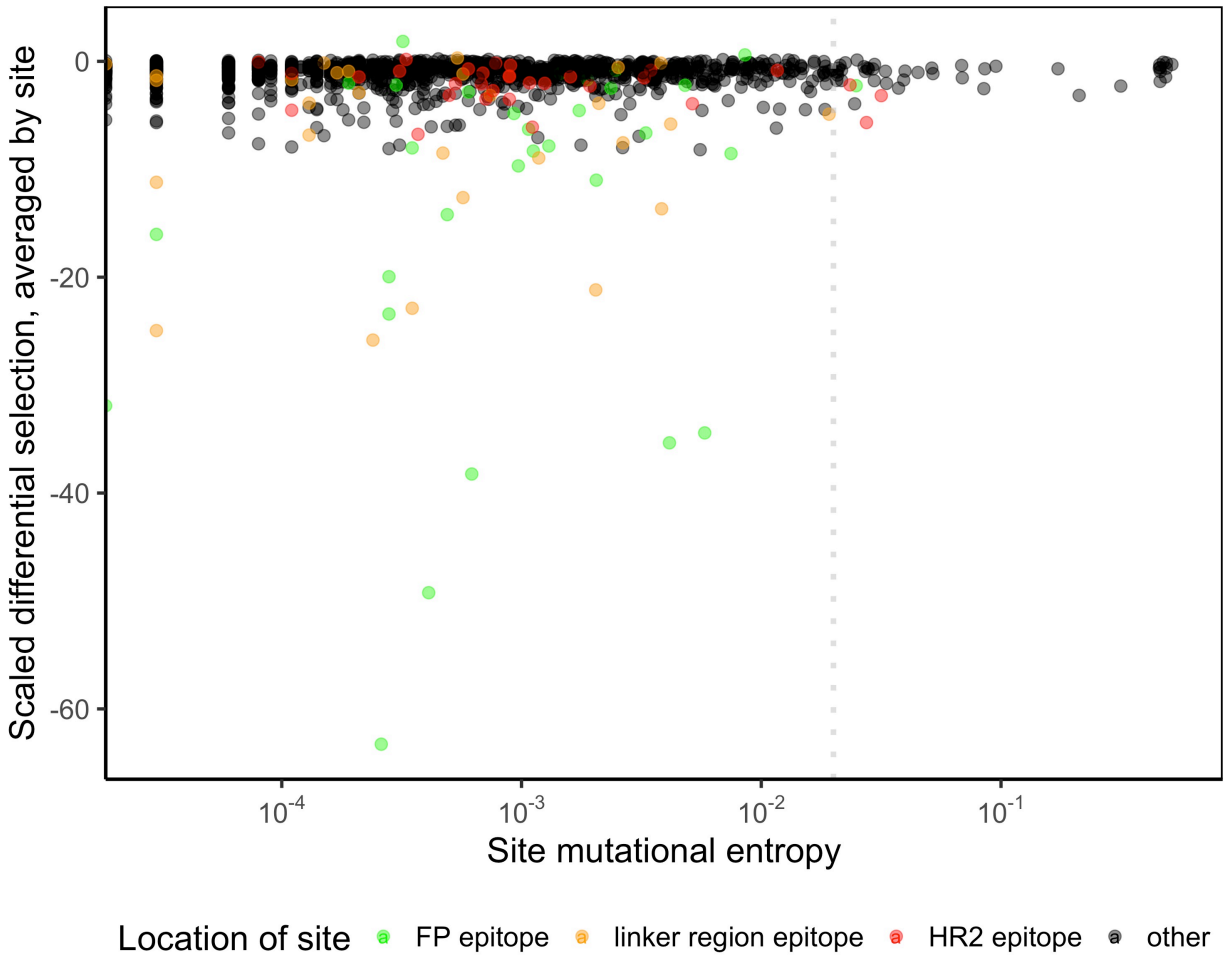
Supplemental Figure 3.3. Plasma binding and neutralization with RBD depleted plasma, related to Figure 3.2. (A and B) Change in plasma binding to RBD (A) and Spike (B) before and after depletion of RBD-binding antibodies, as measured by ELISA area under the curve (AUC). (C) Plasma binding to S2 subunit protein, as measured by ELISA AUC. Dotted lines indicate the lower limit of detection, as determined by pooled pre-pandemic serum. Each point is colored by patient as indicated to the right. (D) Comparison of neutralization titer 50% (NT50) before and after depletion of RBD binding antibodies. The percent residual neutralization for each patient is shown on the right. The dotted line indicates the lower limit of detection (NT50 of 20). Because the lowest dilution of plasma tested is 1:20, we cannot determine NT50 titers smaller than this. (E) Correlation between the most enriched peptide within the FP and HR2 epitope regions and the residual NT50 values for each patient plasma sample. (F) Correlation between plasma binding to S2 and residual NT50 values for each patient. All plasma tested in these assays were from the 30d timepoint.



Supplemental Figure 3.4. Effect of mutations on binding by COVID-19 patient plasma within various regions, related to Figure 3.2. Heatmaps depicting the effect of all mutations, as measured by scaled differential selection, at each site within the NTD, RBD, and CTD regions. Mutations enriched above the wildtype residue are colored blue and mutations depleted as compared to the wildtype residue are colored red. The wildtype residue is indicated with a black dot. Line plots showing the enrichment of wildtype peptides for each patient are shown above, with a solid line for patient samples taken at day 30 p.s.o. and a dashed line for patient samples taken at day 60 p.s.o. Peptides missing from the library are shown as grey boxes in the heatmaps and as breaks in the line plots.

```
                816             833
229E             saiedilfsklvtsglgt      18
NL63             saledllfskvvtsglgt      18
OC43             saiedllfdkvklsdvgf      18
HKU1             slledllfnkvklsdvgf      18
SARS_CoV_2       sfiedllfnkvtladagf      18
SARS_CoV_1       sfiedllfnkvtladagf      18
MERS             saiedllfdkvtiadpgy      18
                 * :**:**.*:  :. *
```

Supplemental Figure 3.5. Multiple sequence alignment of the FP for SARS-CoV-2 and human endemic coronaviruses (OC43, HKU1, NL63, and 229E), related to Figure 3.3. Alignment was performed using Clustal Omega, and amino acids are colored according to physiochemical properties. Red indicates small and hydrophobic molecules (excluding Y), blue indicates acidic molecules, magenta indicates basic molecules (excluding H), and green indicates glycine, hydroxyl, sulfhydryl, and amine molecules. GenBank accession numbers: YP_009724390.1, YP_009555241.1, YP_173238.1, YP_003767.1, and NP_073551.1, respectively.



Supplemental Figure 3.6. Effect of commonly circulating S protein variants on antibody escape for all patients, related to Figure 3.5. Scatterplot comparing the effect of mutations on patient plasma antibody binding and the frequency of all circulating S protein variants. The mutational entropy of every circulating protein variant, as reported at the https://cov.lanl.gov website and based on GISAID global sequencing, is plotted on the x-axis. The average scaled differential selection values for all mutants, averaged across all patients, at each site is plotted on the y-axis. Each site is colored by its location, as indicated on the bottom. The dotted line denotes the cutoff (0.02) of mutational entropy which GISAID uses to determine variants of interest.

SUPPLEMENTAL TABLE

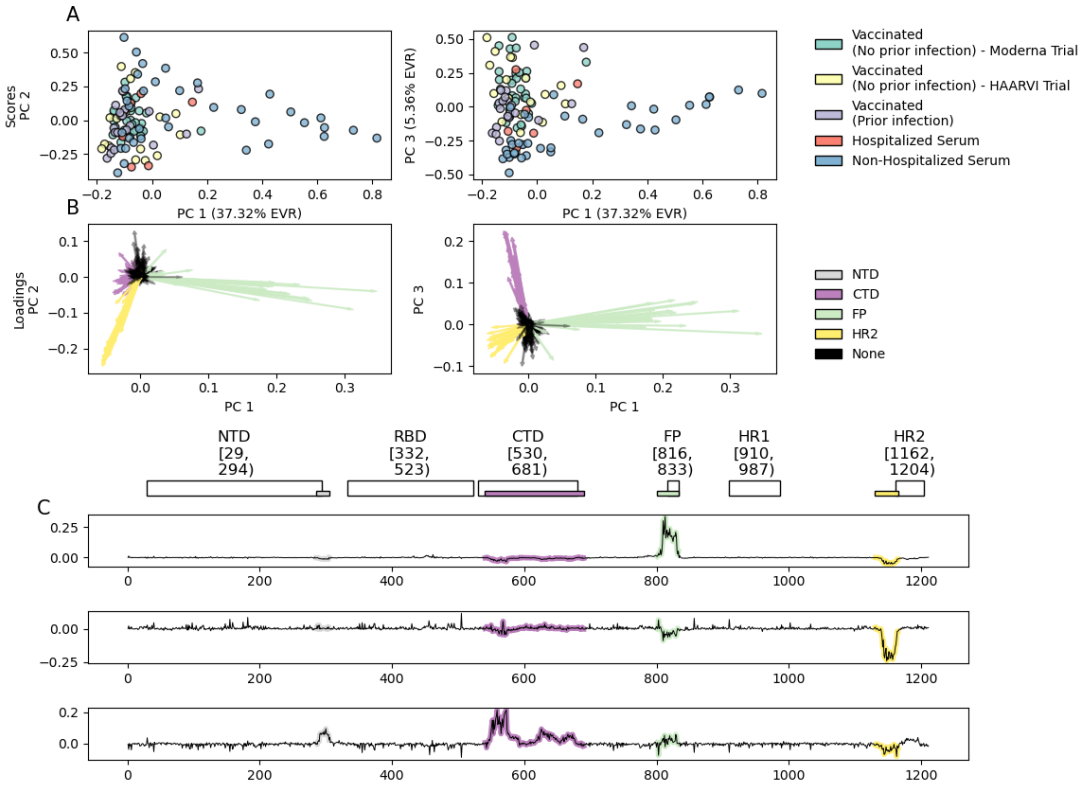
Supplemental Table 3.1. Description of the COVID-19 patient samples used in this study, related to Figure 3.2. All patients exhibited mild symptoms not requiring hospitalization except for patient 6 (indicated by an asterisk), who had moderate symptoms requiring non-invasive ventilation or a high flow O_2 device.

Participant ID	Day(s) post symptom onset (p.s.o.)	Age	Gender
1	27, 72	47	Female
2	31, 60	43	Female
3	29, 71	65	Male
4	31, 67	29	Male
5	31, 63	48	Female
6*	33, 76	64	Female
7	29, 74	22	Male
8	31, 67	31	Female
9	26, 55	56	Male
10	34, 67	28	Female
11	34, 62	30	Male
12	26, 58	36	Female
13	28, 66	65	Male
14	26, 64	65	Female
15	48, 77	52	Female
16	35, 69	52	Male
17	30, 57	36	Male
18	43, 73	29	Male

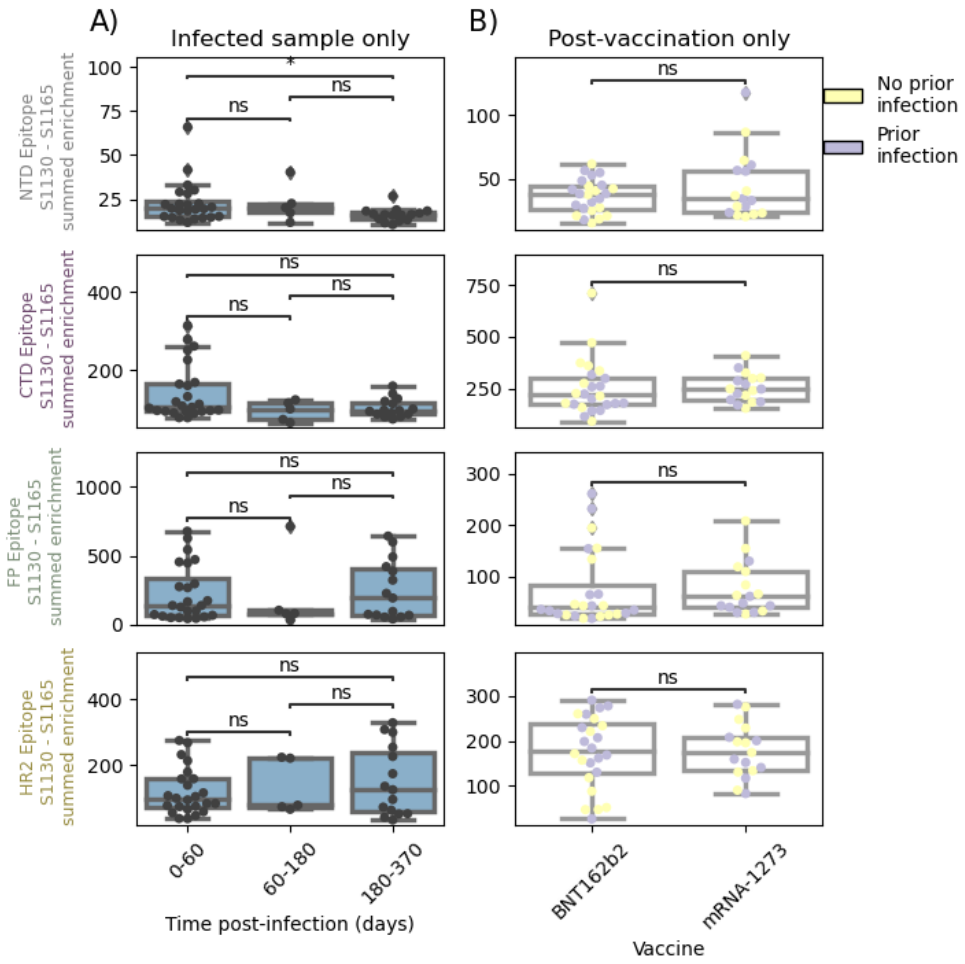
Appendix C

SUPPLEMENTARY FILES FOR CHAPTER 4

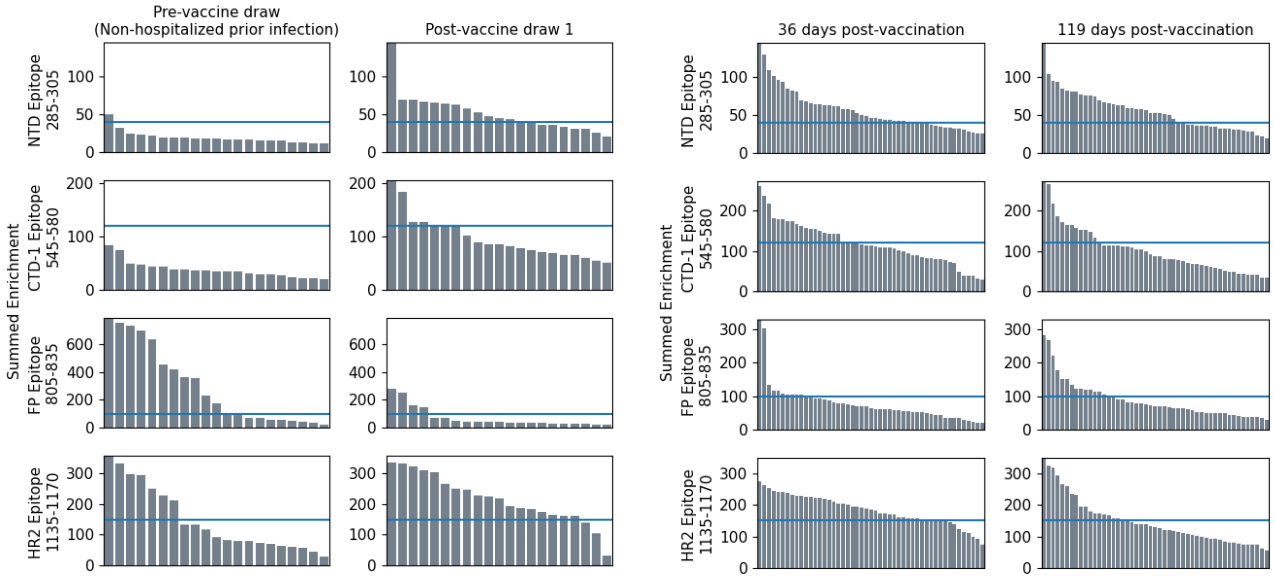
SUPPLEMENTAL FIGURES



SUPPLEMENTAL FIGURE 4.1: Principal Component Analysis on wild-type enrichment features of all samples (A) Scatterplot depicting the unit scaled sample "scores" represented by the columns to visualize sample relationship in principal component space. Colors represent the group which each sample belongs to. (B) Vector plots showing the component loadings, scaled by the square root of the respective eigenvalues in the eigen-decomposition. Colors represent the genomic location of each component loading score. (C) Line plots showing the first three principal axes/directions in feature space, plotted as a function of the wild-type peptide feature location on Spike.



SUPPLEMENTAL FIGURE 4.2: Comparison of epitope binding for HAARVI subgroups. Boxplots of summed wild-type enrichment within epitope binding regions for samples grouped by (A) timepoint post symptom onset or (B) vaccine type (Pfizer/BioNTech BNT162b2 or Moderna mRNA-1273). Results of a Mann-Whitney test between the groups are shown. P-values were adjusted for multiple testing using Bonferroni correction. * indicates $p<0.05$, ns means “not significant”.



SUPPLEMENTAL FIGURE 4.3: Thresholding of total epitope binding within major epitope regions. Histogram showing the summed enrichment values within each epitope region for every sample in the Moderna Trial Cohort (left two panels) or HAARVI Cohort (right two panels). Blue line delineates the threshold chosen for each epitope region. Samples above the line were included in the escape profile analyses.