Understanding Antigenic Variability by Building Novel Deep Mutational Scanning Tools

by

I. M. Francino Urdaniz

B.Sc., Universitat de Barcelona, 2019

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Chemical and Biological Engineering

2023

Committee Members:

Timothy A. Whitehead, Chair

Corrella Detweiler

Jerome M. Fox

Theodore W. Randolph

Kayla G. Sprenger

Francino Urdaniz, I. M. (Ph.D., Chemical Engineering)

Understanding Antigenic Variability by Building Novel Deep Mutational Scanning Tools

Thesis directed by Prof. Timothy A. Whitehead

The immune system is regulated by protein-protein interactions. When a viral infection is detected, the immune system responds by generating antibodies that can neutralize the viral proteins. Meanwhile the virus replicates, generating new strains that could escape such neutralization. Then, the race begins, not only between the virus and the adaptive immune system, but also against the scientists who are developing antibody therapies and vaccines.

For the development of effective therapeutics and vaccine immunogens, the functional sequence space of the binding protein has to be studied. Current methods rely on the study of existing variants of a viral glycoprotein. However, with every emerging variant, new therapeutics and vaccine immunogens have to be developed. In this thesis I have used deep mutational scanning coupled with next generation sequencing to build tools that will contribute to prospectively map the surface of a protein. I hypothesize that being able to identify escape mutants on the developed therapies as well as characterizing the tolerable sequence variation of the targeted protein will contribute to the generation of new and more potent therapeutics and vaccine immunogens.

Dedication

I dedicate this work to my family, who has taught me to take every opportunity as they come and make the most out of them.

Acknowledgements

I want to express my gratitude to my advisor, Prof. Tim Whitehead, for helping me become the scientist I am today. His evolving mentorship as I grew as a researcher and his encouragement to follow my research ideas made me want to pursue a PhD and have made this thesis possible. Thank you for all the experience I have gained through my graduate years.

I thank the members of my committee for their support and suggestions to make this thesis better and more complete. And to the past and present members of the Whitehead lab, they have been an amazing team. Thank you for the moral support and scientific discussions that have helped me through my projects. I would also want to express my gratefulness to the former Balsells foundation, their fellowship made this adventure possible from the start.

Finally, I am extremely grateful for the friends I have made along the way who have become my Colorado chosen family. Thank you for your emotional support and for making this time a memorable experience. Most importantly to my family, despite being an ocean away, they have been a constant support and encouragement.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction

1.1   Purpose of the study

Protein-protein interactions (PPI) regulate fundamental biological processes including the immune system and some viral infections. A general mechanism of viral infection is through the interaction of a surface protein of the virus with a surface cell protein of the host. When the infection is detected, the body generates antibodies that will bind to the viral protein and signal it to be eliminated from the host. Most of the time this binding happens on the footprint of the host cell, preventing the virus from recognizing the host and thus neutralizing the viral infection.

Mimicking the immune system, scientist are developing therapeutics, using antibodies as well as small molecules, that also block that interaction from happening. In order to generate this kind of therapeutics, the surface footprint between the two proteins, known as the epitope, has to be studied and a well-established technique to do so is through epitope mapping. The mapping of protein interactions can happen through different techniques and all give different and complementary information. With the emergence of SARS-CoV-2 in 2019, there was a huge effort to map the interaction of the receptor binding domain (RBD) with the host cell receptor ACE2. All of the possible techniques were used in parallel to analyze this new protein and map were the antibodies were binding. A more in depth review of the epitope mapping techniques used in the study of SARS-CoV-2 S RBD have been used is given in **Chapter 2**. In it, I compare the

information obtained with each technique and we put a special focus on deep mutational scanning since, to our knowledge, there is no previous review [1].

When the viral protein is known, therapeutics can be designed to neutralize it. However, these are designed against a specific strain of the virus. This means that when a new strain emerges, the neutralization power of this antibody could be compromised or even completely lost. The mutations that escape the neutralization are called escape mutants. When the escape mutants escape neutralization of the existing therapeutics, the process has to be restarted to generate new therapeutics that can neutralize the current strain of the virus. This renders the process inefficient making the study of protein-protein interactions and the tolerated sequence space a research focus.

1.2   Scope of the study

I hypothesize that deep mutational scanning can be used to build tools that will facilitate the study of viral proteins. By understanding what mutations are tolerated by the viral protein and which ones escape antibody neutralization, we can inform the design of new therapeutics. In this thesis, I have used deep mutational scanning in the context of SARS-CoV-2 S RBD to build widely applicable tools for understanding the viral protein antigenic variability. In particular, I study the Spike protein Receptor Binding Domain (S RBD) which is responsible for binding its host receptor ACE2. Currently, any SARS-CoV-2 variant must bind the ACE2 receptor to enter the host cells, thus, any future VOC will most likely need to maintain this binding.

First, I displayed the Wuhan-1 S RBD on the surface of yeast, verified that the protein was folding correctly and that we could obtain reproducible results. With it, I characterized the epitope of the monoclonal antibody 910-30 [2]. Then, using the displayed S RBD, I developed a tool for identifying mutations on this protein that would escape antibody neutralization while still maintaining binding to ACE2[3]. In the study, I identified single point mutations that escape neutralization of antibodies isolated from convalescent patients [4]. Some of the identified mutants later emerged in variants of concern such as K417N/T, N501Y or Q493R (**Chapter 3**). Next, I

used this tool to inform the engineering of broadly neutralizing antibodies in the context of the Omicron VOC [5]. Knowing when the antibodies would fail to neutralize, our collaborators in Jardine's lab did several rounds of selection to engineer antibodies that not only neutralized Beta and Gamma strains, but also Omicron (**Chapter 4**).

Finally, I wanted to better understand the mutational space on the RBD with highly mutagenized libraries. For this, I benchmarked a cassette-based method using Golden Gate that allows for building custom, user-defined libraries on the order of 10^4 to 10^7 unique protein encoding variants [6]. A limitation of deep mutagenesis is the generation of the libraries. Usually the mutagenesis methods are time consuming, the final library contains considerable amounts of WT and the final sequences not always match exactly the initial design. Using customized oligos and Golden Gate assembly, in a single day, I generated libraries containing less than 2% WT and containing all the designed mutations (**Chapter 5**). I used this method to build large customized libraries and train a computational model to predict the effect of the combination of different mutations on ACE2 binding (Manuscript in preparation). The built libraries cover a sequence space far away from current variants of concern and interest that, to our knowledge, has not been previously studied. I have identified a low mutational space on the surface of S RBD that could be used as target for new therapeutics (**Chapter 6**).

1.3 Limitations

Yeast surface display is widely used for understanding protein-protein interaction as well as for engineering proteins. It has many advantages like the incorporation of post-translational modifications, the inclusion of epitope tags and the compatibility with flow cytometric analysis. Nonetheless, it also has some limitations regarding the size of the library that can be screened, a different glycosylation pattern than that of human cells or the limitation of the protein size that can be properly expressed and folded[7]. In this studies, a potential limitation of using yeast surface display as the main platform is the accurate folding of the RBD compared to the packaging of this

protein in the virus within the whole Spike. To overcome this limitation, I have used pseudovirus analysis when convenient to reproduce the behaviour of the RBD libraries. The results have been in agreement with the yeast surface display.

While the RBD of the S protein is responsible for recognizing the ACE2 host protein, it is not the sole protein on the surface of the virus. Other proteins as well as different parts of the Spike protein play an important role on cell entry and infection. [8]. This is another limitation of using yeast surface display is the study of RBD by itself. Using pseudovirus assay, the behaviour of the generated yeast plasmid libraries can be studied in the context of full virus.

I have used the S RBD to build deep mutational tools that I believe are widely applicable to other protein-protein interactions. However, I have not had the chance to tested them on a different protein. To ensure the applicability of these tools I have used general plasmid backbones in which any protein can be encoded and generalized the protocols.

Chapter 2

An overview of methods for the structural and functional mapping of epitopes recognized by anti-SARS-CoV-2 antibodies

This mini-review presents a critical survey of techniques used for epitope mapping on the SARS-CoV-2 Spike protein. The sequence and structures for common neutralizing and non-neutralizing epitopes on the Spike protein are described as determined by X-ray crystallography, electron microscopy and linear peptide epitope mapping, among other methods. An additional focus of this mini-review is an analytical appraisal of different deep mutational scanning workflows for conformational epitope mapping and identification of mutants on the Spike protein which escape antibody neutralization. Such a focus is necessary as a critical review of deep mutational scanning for conformational epitope mapping has not been published. A perspective is presented on the use of different epitope determination methods for development of broadly potent antibody therapies and vaccines against SARS-CoV-2.

2.1   Introduction

The interaction of proteins with other proteins is foundational to cellular life [9]. Understanding the structural, functional, and mechanistic basis of such noncovalent protein-protein interactions can help rationalize emergent cellular behavior [9], can be exploited for design of biologics like antibodies [10] and can also be used to map and predict the next moves in the trench warfare between humoral immunity and pathogen evasion and evolution [3].

An important class of protein-protein interactions are antibody interactions with antigens. Here, the epitope is defined as the antigenic surface recognized by a given antibody. Identifying the structures, sequences, and sequence constraints on such antigen epitopes is essential for solving difficult problems in basic and applied immunology. For example, a key idea in modern vaccine design has been that antigen structures can be modified rationally to present critical epitopes that elicit antibodies that neutralize infection (neutralizing antibodies or nAbs) that, in turn, confer long-lasting protection. The first proof of concept demonstration of such structure-based vaccine design in Phase I clinical trials was published [11] for an immunogen mimicking a key conformational epitope of a viral protein in respiratory syncytial virus. Similarly, the search for a universal influenza A vaccine was jump-started by the structural and sequence identification of a conserved epitope on the influenza surface protein haemagglutinin [12, 13, 14]. Antibodies targeting this haemagglutinin epitope are able to neutralize broadly across different influenza A subtypes. This structural definition of an epitope led to immunogen designs that elicit high levels of broadly neutralizing antibody titers in a recently completed phase I clinical trial [15]. Thus, therapeutic and prophylactic strategies are informed by, and often start with, a sequence and structural definition of an antigenic epitope.

There exist several relatively mature technologies available to delineate the sequences, structures, or sequence constraints of epitopes. In fact, several comprehensive reviews of individual methods have been published in this century [16, 17, 18, 19, 20, 21, 22, 23]. **Table 2.1** lists com-

mon experimental methods for epitope mapping. There are two major classifications of epitopes primarily based on the experimental method used for their identification. Linear epitopes are those that involve sequential residues in the primary amino acid sequence and can be identified using techniques like peptide microarrays, phage, or bacterial display. By contrast, conformational epitopes involve surfaces recognized by antibodies only when a protein is folded in its tertiary or quaternary state. Such conformationally sensitive epitopes are typically resolved by structural determination using X-ray crystallography or electron microscopy (EM). Less commonly, hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) [23] or deep mutational scanning [24] can be employed. All methods have their relative strengths and drawbacks, but generally it has been difficult to compare directly between methods as not all are typically performed on the same set of proteins.

The emergence of SARS-CoV-2 [25] has led to intense research on its virology, epidemiology, and therapeutic and prophylactic interventions [26]. During this time, dozens of research groups around the world identified antibodies raised against natural SARS-CoV-2 infection [2, 27, 4, 28, 29]. This outpouring of research represents a natural experiment for the relative strengths, weaknesses, and types of information inherent in different epitope mapping methods. Thus, in this review we critically survey techniques used for epitope mapping on SARS-CoV-2. However, we do not intend an in-depth explanation of all the methods since exhaustive modern reviews already exist and are cited. Nonetheless, an additional focus on this mini-review is given on epitope mapping and identification of mutants which escape antibody neutralization using deep mutational scanning [24], as to our knowledge no comprehensive review exists. Thus, the second half of this review is given to the critical appraisal of different deep mutational scanning strategies because since the effect of individual mutations on binding can be studied, deep mutational scanning is especially relevant when developing antibodies against evolving viruses.

Given that well over a hundred thousand papers have been published on SARS-CoV-2 [26], a comprehensive review is impractical for this short mini-review format. We apologize to colleagues

Table 2.1: **Summary of common experimental methods for linear and conformational epitope mapping**

Category	Technique	Information Obtained	Comparative Advantage	Comprehensive review
Linear epitopes	Peptide Arrays	Linear peptide sequence recognized by antibody	Massive parallelization allows proteome-size scalability	Katz el al. [19]
	Phage and Bacterial Display		Can use linear and constrained peptides in a high throughput format	Pande et al.[21]
Conformational epitopes	Electron Microscopy (Cryo-/negative stain)	Atomic structure of an antigen-antibody complex	Structural determination of large, complex complexes with only small amounts of material needed	Renaud et al. [18]
	X-ray Crystallography		Highest quality atomic structural determination	Malito et al. [16]
	HDX-MS	Antigenic surfaces shielded from solvent in presence of antibody	Description of dynamic conformations	Sun et al. [23]
	Deep Mutational Scanning	Comprehensive antigenic sequence determinants to binding /competitive inhibition	High resolution sequence constraints on antigenic epitopes and evaluation of point mutants	

whose work we have failed to cite.

2.2   SARS-CoV-2 Spike protein as a model system

Comparisons between epitope mapping methods can be accomplished on antibodies targeting the same protein. In this mini-review we will focus on the SARS-CoV-2 Spike protein (S) as it is a highly glycosylated surface exposed protein on the virus and the focus of the overwhelming majority of SARS-CoV-2 epitope studies published to date **(Figure 2.1 A)** [4, 30, 31]. The S protein is a homotrimer in which each protomer is arranged as two subunits, S1 and S2. A furin cleavage site separates each S subunit and after cleavage the subunits are noncovalently associated in the prefusion metastable structure [32]. The S1 subunit binds to angiotensin-converting enzyme 2 (ACE2) via its receptor binding domain (RBD) [33] and contains an N-terminal domain (NTD), while the S2 subunit containing the C-terminal domain (CTD) is critical for the fusion of the viral and host cell membranes. The S2 subunit is more conserved than S1, perhaps because most of the surface exposed portion of the virus is on S1 [34]. Similar to other coronaviruses, the prefusion metastable structure of S undergoes two major conformations: a conformation where the RBD is in the "up" state and a conformation with RBD "down" [28, 35]. The biological relevance for these conformations is that the ACE2 receptor binding motif (RBM) is exposed to solvent only when the RBD is in the "up" state. Thus, at least one RBD must be in the "up" state for cell entry via ACE2 recognition.

2.3   X-ray crystallography

X-ray crystallography allows atomic resolution of the antigen-antibody interaction and is the acknowledged gold standard for epitope determination. Epitopes can be determined amazingly fast: the first structure of an anti-SARS-CoV-2 antibody in complex with S RBD [28] was reported on a preprint server only 9 weeks after the genetic sequences of SARS-CoV-2 were made public. Closely following this initial study, other reports described neutralizing and non-neutralizing epitopes for antibody complexes with individual S domains like the RBD **(Figure 2.1 B)** or NTD **(Figure 2.1**
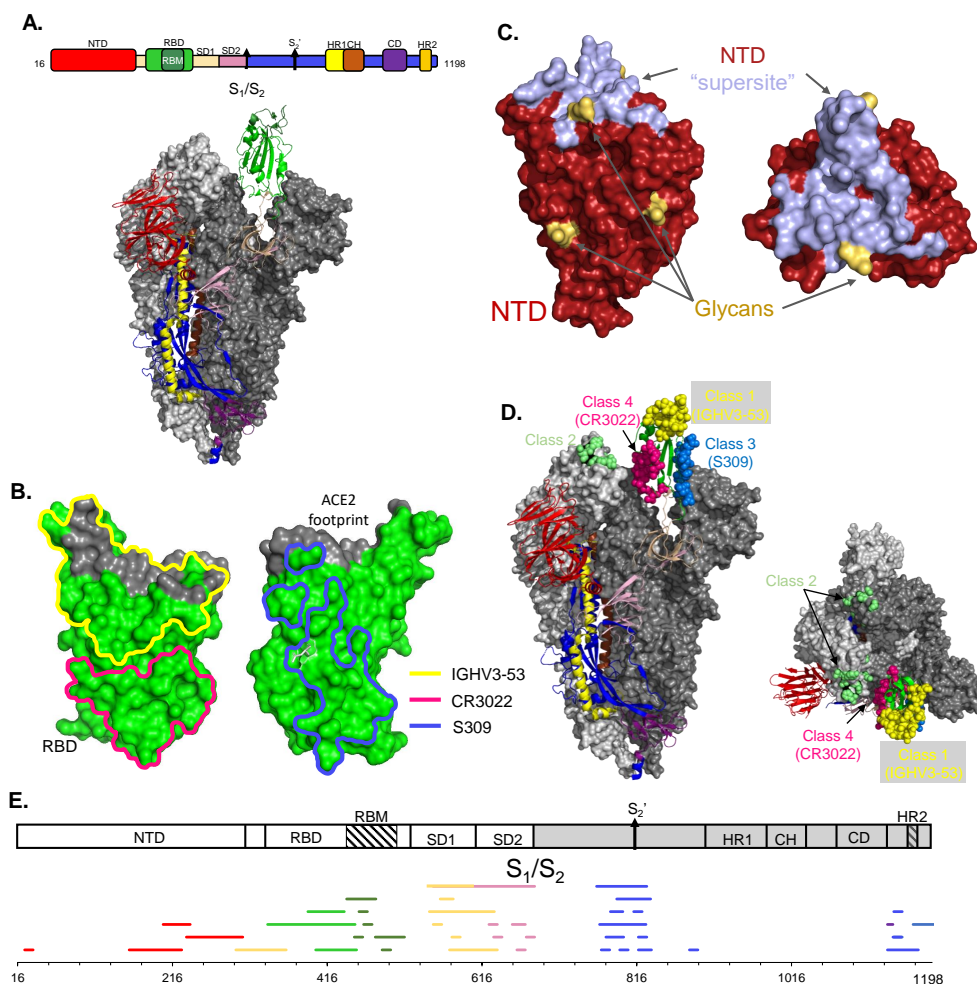
Figure 2.1: **Epitope mapping techniques in the context of SARS-CoV-2. A.** SARS-CoV-2 Spike ectodomain schematic with labelled regions. NTD: N-terminal domain, RBD: receptor-binding domain, RBM: receptor-binding motif, SD1: subdomain 1, SD2: subdomain 2, S1/S2: furin cleavage site, S2': S2' cleavage site, HR1: heptad repeat 1, CH: central helix, CD: connector domain, HR2: heptad repeat 2. Below is the structure of the Spike glycoprotein color coded with one protomer shown with RBD in the up conformation. The other two protomers are shown in different shades of grey and have the RBD in the down position. **B.** RBD structure (in green) showing epitopes identified by x-ray crystallography for anti-RBD IGHV3-53 (yellow), CR3022 (pink), and S309 (blue) antibodies. **C.** NTD structure (in chocolate brown) shown with the "supersite" epitope (pale blue). Glycans are shown in yellow. **D** Common epitopes represented as spheres for the identified antibody classes on S. Class 1 binds on the RBM only available on the 'up' conformation. Class 2 can recognize the RBD on the 'up' and 'down' position. Class 3 binds in the same region as the previously identified S309 nAb. Class 4 in a non-neutralizing group of antibodies that bind a conserved epitope only available on the 'up' conformation, previously identified with CR3022. **E.** Linear epitope mapping along the Spike protein. Epitopes are color coded by domain as in a. Note the diversity of epitopes, including at SD1 and SD2 domains that are underrepresented in X-ray crystallography and EM structural studies.

C) [35, 36, 37, 38, 39, 40, 41]. These early studies helped define the structural basis of neutralizing and non-neutralizing epitopes on these individual domains.

The RBD is a major target for neutralizing antibodies since it is responsible for binding ACE2 [32]. In the early days of the COVID-19 pandemic, antibodies from SARS-CoV convalescent patients were screened against SARS-CoV-2 S RBD. An early cross-reactive antibody is CR3022 [28], and this antibody defines one non-neutralizing and broadly conserved epitope on RBD distal to its RBM **(Figure 2.1 B)**. Another early described broadly conserved epitope is the one recognized by mAb S309 [36] **(Figure 2.1 B)**, which recognizes an epitope defined by a conserved N-linked glycan at Asn343. In contrast to the CR3022 epitope, antibodies at this S309 epitope neutralize both SARS-CoV and SARS-CoV-2. Further into the pandemic, SARS-CoV-2 specific nAbs were identified from convalescent patients and, for some, their epitopes were structurally determined by X-ray crystallography. Some examples are P2B-2F6 [42], P4A1 [43], CB6 [44]; some other antibodies such as PR1077 [45] were isolated from immunized mice. A large fraction of these nAbs bind at or adjacent to the ACE2 binding site. In particular, P4A1 [43] covers the majority of the ACE2 footprint. As one example, nAbs from the IGHV3-53 germline class represent the most common antibodies elicited from natural infection from the original Wuhan-Hu-1 strain [37]. Structures of IGHV3-53 nAbs CC12.1, CC12.3, and B38 define the basis of neutralization by competitive inhibition of ACE2 recognition [37, 38] **(Figure 2.1 B)**.

Antibodies can also neutralize SARS-CoV-2 by binding at the NTD, with several crystallographic studies pinpointing the key epitopes [39, 40]. There are conserved epitopes between SARS-CoV and SARS-CoV-2 NTD but all are non-neutralizing; conversely, the key non-conserved epitope is neutralizing and has been named 'supersite' [39, 41] **(Figure 2.1 C)**. Most of the NTD surface is covered by a glycan shield, and the supersite is one of the only exposed proteinaceous surfaces on NTD. Structural studies show that antibodies from different germline classes bind this key aglycosylated epitope [40]. Unfortunately, this supersite undergoes extensive antigenic variation, and many variants of concern (VoC) are no longer neutralized using supersite nAbs elicited

from the original Wuhan-Hu-1 strain [39].

Overall, X-ray crystallography has been a key technique in the study of SARS-CoV-2 epitopes as it was used to define individual conserved and non-conserved epitopes on the RBD and NTD of the SARS-CoV-2 S. Key limitations of this technique include the difficulty of the preparation of high diffraction quality crystal of full-length S ectodomain, limiting determination of epitopes to those entirely contained within individual RBD and NTD domains.

2.4   Electron Microscopy

The 2017 Nobel Prize in Chemistry was awarded "for developing cryo-electron microscopy [cryo-EM] for the high-resolution structure determination of biomolecules in solution". The use of electron microscopy, and cryo-EM in particular, has exploded in popularity over the past decade for the method's ability to determine structures of large protein complexes like antibodies in complex with S. In fact, cryo-EM was used to determine the atomic structure of S ectodomain less than two months after the publication of the S sequence [32, 35].

Dozens of cryo-EM and, less commonly, negative stain-EM structures [46] of potent neutralizing antibodies in complex with S have been reported. We list here a few of the antibodies that can be grouped in two representative examples of the types of epitopes that can be analyzed using electron microscopy. In the first example, a study led by Adimab scientists used cryo-EM to determine the epitope of a broadly neutralizing antibody developed by Adimab that binds to S RBD [47]. Regeneron has also developed an antibody cocktail binding to S RBD whose epitope has also been mapped using this same technique [48]. These specific complexes could have also been determined by X-ray crystallography since the epitope is entirely contained within an S RBD monomer. Novel epitopes such as the one of H014 on RBD [48] and the anti-S NTD antibody 4A8 [41] can also be determined using EM. In another example, a different group used cryo-EM to characterize the epitope for a nAb that binds simultaneously to two of the three RBDs contained in the S trimeric complex [49]. This specific complex would be difficult to determine by X-ray

crystallography. Thus, cryo-EM can be used for complexes both amenable and refractory to solution by X-ray crystallography. Cryo-EM and X-ray crystallography can be combined to define the structural epitopes recognized by antibodies elicited from natural infection. An excellent example of a joint study was reported by Barnes et. al, who defined the four major classes of antibodies binding to RBD epitopes [50] (see definitions in **Figure 2.1 D)**.

2.5   Linear Epitopes from Synthetic Peptide Arrays

Linear epitopes are commonly identified using synthetic peptide arrays [19, 20] or peptide libraries coupled with phage or bacterial display [21, 22]. Most of the published linear epitopes for SARS-CoV-2 have been from synthetic peptide arrays; to our knowledge there have been no published reports on the use of phage display to determine epitopes on SARS-CoV-2. A unique strength about determining epitopes using synthetic peptide arrays, compared with other techniques covered in this review, is that individual antibodies as well as a bulk serological response can be studied.

Synthetic peptide arrays have been used to study epitopes of monoclonal antibodies and convalescent patient serum on the whole S protein by several groups [51, 52, 53, 54, 55, 56]. Even though this review focuses on the S protein epitope mapping, one group has used synthetic peptide arrays to identify proteome-wide epitopes for SARS-CoV-2 and other coronaviruses [51], highlighting the advantage of scale for synthetic peptide arrays.

The identified linear epitopes on S are clustered in defined regions (**Fig. 2.1E**) mainly at cleavage sites or sites necessary for conformational changes for viral entry, like the S1/S2 cleavage site [37], the S2' cleavage site, and the CTD [52, 53]. While the majority of the linear epitopes are found outside of the RBD, several have also been identified on the RBD [54]. These combined studies highlight the diversity of the antibody response on the entire S protein and pinpoint immunodominant epitopes as well as epitopes that are relatively occluded from antibody recognition. However, there is a lack of information on the correlates of protection for these identified epitopes,

and the structural basis for recognition must be inferred by structural information given by cryo-EM and X-ray crystallography.

2.6 Hydrogen Deuterium exchange

Epitope determination using hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) is based on the biophysical principle that amide hydrogens can exchange with deuterium in deuterated solvent faster when unbound than bound with antibody. Epitopes are determined by identifying locations on protein surfaces with lower exchange rates. The S RBD-ACE2 interface as well as the soluble ACE2 have been mapped by HDX-MS, which contributed to our understanding of the conformational changes on the S protein upon binding to ACE2 [57]. HDX-MS has also been used to determine and explain antibody epitopes [58, 59]. Regeneron in particular used HDX to understand the mechanism by which non-competitive antibodies bind simultaneously to the RBD [57]. These results can help create a cocktail of neutralizing antibodies that would not overlap or block each other while simultaneously binding the RBD on the ACE2 footprint. While HDX-MS can facilitate the understanding of the conformational dynamics of binding, it may give recurrent false positives and the experimental proposal must fulfill an exacting list of requirements to obtain good results [23]. Thus, HDX-MS is usually coupled to methods like cryo-EM to marry conformational dynamics with structural insight.

2.7 Deep mutational scanning

Deep mutational scanning, independently developed by the Fields [24] and the Bolon [60] groups, leverages next generation sequencing to observe the functional effect of individual mutants in a large population [24]. The power of this method is scale, as tens of thousands of mutants can be assessed in a mixed pool. In 2015, conformational epitope mapping of protein-protein interactions using deep mutational scanning was independently developed by different labs [61, 62, 63]. In the last year, three independent groups have used similar epitope mapping approaches to understand and engineer interactions between S RBD and antibodies or the ACE2 receptor **(Figure 2.2)**. The

Procko group identified mutations on human ACE2 that increase binding affinity to S RBD [64]. The Bloom group identified the sequence determinants of S RBD for ACE2 recognition and mapped anti-RBD antibody epitopes [31, 65, 66]. Finally, the Whitehead group has developed a method to determine the set of mutations on S RBD which can escape monoclonal antibody neutralization [3].

In deep mutational scanning, the antigenic sequence dependence on binding can be assessed for nearly every single point mutant in the protein sequence. This information is used to identify conformational epitopes under the assumption that epitope positions are less tolerant of mutations than non-epitope positions. Deep mutational scanning workflows for conformational epitope mapping are similar at a superficial level. The antigen of choice is displayed on the surface of a eukaryotic cell. Next, binding to an antibody or receptor is monitored using a flow cytometer after cell labeling with appropriate fluorophores. Comprehensive mutagenesis of the antigen gene is performed thus generating a library of antigen mutants that can be transformed into the relevant cell type. A population of cells, where each cell displays a distinct antigen mutant, is split and incubated in several different binding conditions. For example, each reaction could contain a different amount (or none) of antibody. After fluorophore labeling, the cells are screened using a cell sorter. Different populations are distinguished using gates on different light scattering or fluorescent values. For example, a gate is typically set to identify cells maintaining high antibody binding as inferred by a high fluorescence signal in the appropriate channel. Populations of cells are sorted according to these gates by fluorescence activated cell sorting (FACS), regrown, plasmids harvested and prepared for deep sequencing, and then sequenced. For each sorted population the frequency of each variant is enumerated; along with other information about sorting conditions, this information is processed either qualitatively or quantitatively to identify the effect of each introduced mutation on the binding considered in the assay.

The Procko group used deep mutational scanning to identify ACE2 mutations that increase binding to SARS-CoV-2 S RBD [64] in order to develop a receptor trap prophylactic and therapeutic
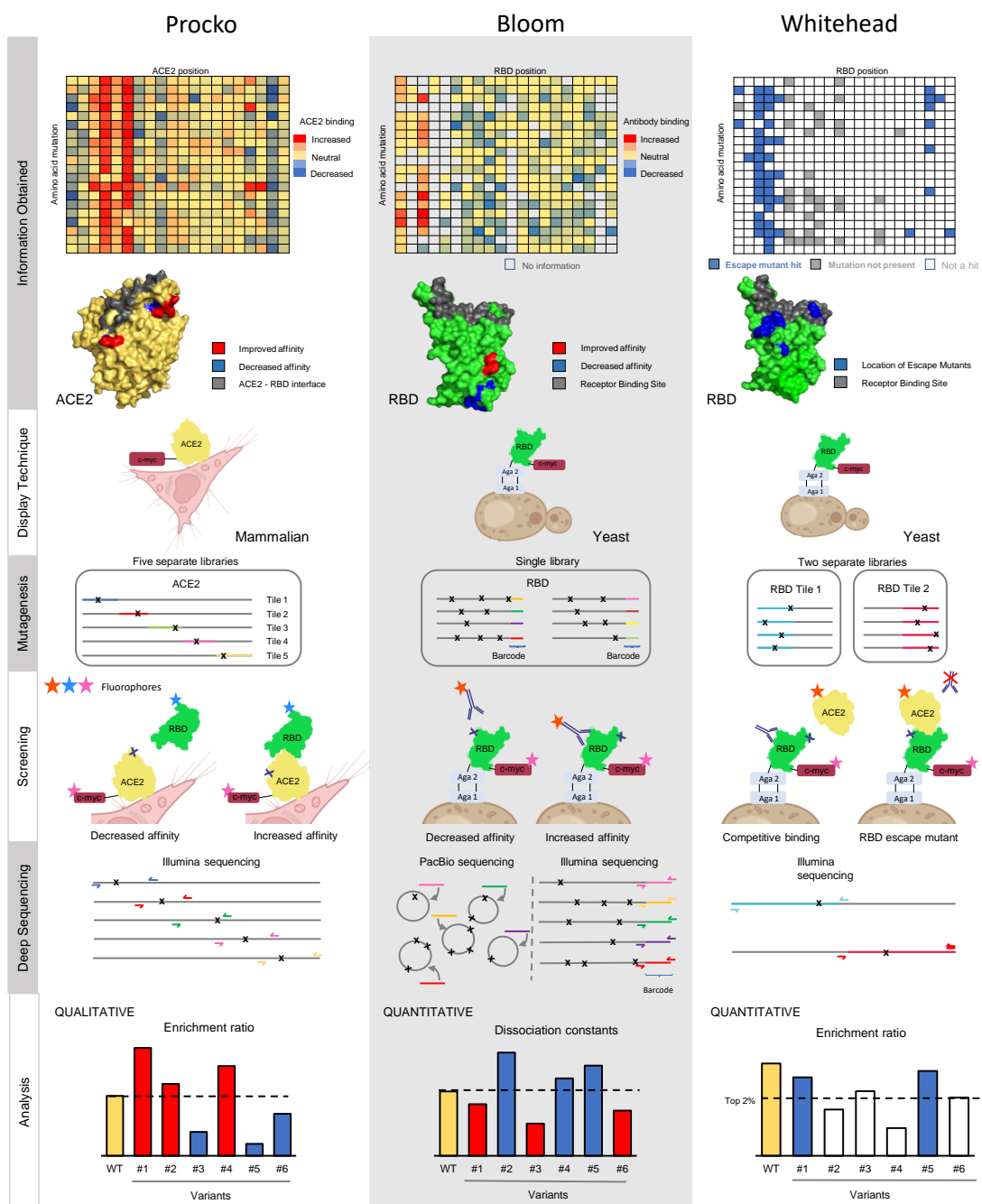
Figure 2.2: **Overview of independent deep mutational scanning workflows for conformational epitope mapping.**

against SARS-CoV-2. Key mutations found to increase ACE2 binding to S were those removing N-linked glycans that partially shield the ACE2 surface recognized by the S RBD. The best engineered soluble ACE2 (sACE2) variant can outcompete natural ACE2 for binding to S RBD. Further, the authors showed that sACE2 can neutralize different coronaviruses, including SARS-CoV and SARS-CoV-2 [67]. To engineer this receptor, ACE2 was displayed on the surface of mammalian cells and incubated with soluble S RBD. The variants that bind tighter to the S RBD than native ACE2 were collected and identified by an increase in frequency in the binding population relative to a control [60].

The Bloom research group used deep mutational scanning for the quantitative assessment of the sequence dependence of S RBD on ACE2 binding affinity [65]. This same platform was also used to map epitopes and escape mutants for several monoclonal antibodies [31, 66], predicting in advance the N501Y mutation observed in several VoC. S RBD is displayed on the surface of yeast and labeled either with soluble ACE2 or mAb at multiple different concentrations. Cell populations collected depend on whether epitopes or escape mutants are identified, and sequence data is processed using a quantitative maximum likelihood estimation method [68].

The Whitehead group has developed a method that identifies the near-comprehensive set of escape mutants on S RBD for neutralizing antibodies that directly compete with ACE2 for binding [3]. Several antibodies can be tested in parallel. Most escape mutations identified in the study are located adjacent to but not directly on the ACE2 binding footprint. Most intriguing, many escape mutants map to K417, including K417N which is present in the circulating Delta plus VoC (B.1.617.2+K417N) and in the Beta VoC and K417T present in the Gamma VoC [69, 70]. To identify escape mutants, an aglycosylated S RBD construct is displayed on the surface of yeast and a competitive binding experiment is performed between a given antibody and soluble ACE2. Cells harboring RBD variants able to maintain ACE2 binding in the presence of a nAb are collected, and a novel algorithm is used to identify escape mutants.

The above studies all performed different strategies, shown in Figure 2, with these differences instructive for those setting up a deep mutational scanning experiment. One major difference between groups is the display technique. One group displayed bona fide ACE2, including its membrane-spanning pass, on mammalian cells, while the other groups used an artificial genetic fusion of S RBD to a yeast cell surface protein. The yeast display set-up maintains several advantages for deep mutational scanning: relatively fast growth rates, excellent genetics and high transformation efficiency, robust cells, and validated protocols [71]. In our hands 11 of 12 tested antibodies targeting S RBD maintained binding to the engineered construct on yeast [3], attesting to the fidelity of the platform. Still, it remains difficult to display complicated glycoproteins in the active form [3], and yeast has different N-linked glycosylation patterns involving heavy mannosylation relative to mammalian cells [72]. Therefore, antibodies that target across S protomers, that involve glycan recognition, or that bind on the S2 protein cannot be considered using yeast display. While mammalian cell display has several disadvantages relative to yeast display, the key advantage is displaying a membrane protein in its native context. In the Procko case, using the native ACE2 conformation was essential to identify that the removal of the glycans increases the binding affinity to RBD.

The two next steps in deep mutational scanning are (i.) performing comprehensive mutagenesis of the gene to be scanned; and (ii.) transforming the resulting DNA libraries into cells. Comprehensive mutagenesis on plasmid DNA can be performed using several methods like PFunkel [73], nicking [74], or overlap extension PCR mutagenesis [64]. Illumina sequencing platforms typically utilize 250 base pair DNA sequencing, which limits the linear stretch of the gene which contains mutations to typical 250-350 bp. Covering an entire gene like ACE2 or S RBD, which are both larger than 350 bp, requires multiple libraries for coverage. These libraries are colloquially referred to as 'tiles'. Both the Procko and Whitehead groups used this tiling strategy (**Figure 2.2**). The main disadvantage of tiling is handling each library independently - separate labeling, sorting, and DNA prep steps must be performed for each tile. In contrast, the Bloom group en-

coded all mutations on S RBD in a single library. Then, they utilized PacBio long read sequencing to haplotype each set of mutants on S RBD to a unique barcode **(Figure 2.2)**. Illumina short read sequencing of the short barcode could then be used to identify frequencies of each mutant. This approach has a higher upfront cost of library haplotyping (the PacBio step) but has more streamlined downstream steps with less expensive sequencing on the backend.

All groups used FACS to screen cell populations. Both Procko and Bloom groups used direct labeling either with antigen or antibody. In contrast, the Whitehead group developed a competitive ACE2 binding screening assay for a neutralizing antibody to infer the set of escape mutants. All groups also used Illumina for next generation sequencing of library DNA. The Procko and Whitehead groups screened and sequenced each tile separately, while the Bloom group sequenced library barcodes only. Best practices for these screening steps involve making true biological replicates (DNA libraries prepared and transformed independently) and sorting replicate libraries on different days.

In the final step, sequencing results are analyzed with a method appropriate for each approach. The analysis results are qualitative or quantitative and depend on factors in the experimental approach like the choice of display format, the type of mutagenesis performed, and screening strategy. In deep mutational scanning workflows the first step is to enumerate the frequency of each variant for each sequenced population. The simplest qualitative analysis is to compare the frequency of a selected population with a reference population that has passed through the cell sorter but is otherwise not screened for binding. The log transform of this frequency change between populations is called an 'enrichment ratio'. The Procko group used this qualitative analysis to determine the relative binding for their ACE2 variants. Such qualitative analyses are simple to perform and suitable for engineering goals like developing superior ACE2 receptor traps. However, this enrichment ratio analysis is subject to consider noise resulting from complexity bottlenecks in the FACS screening, DNA preparation, and sequencing steps. Thus, one drawback from a qualitative analysis is hit identification - how does one determine high enrichment ratios that result from binding events

rather than ones that occur by chance? The Whitehead group solved this problem by independently sorting a control population subject to the same screening criteria as their competitively inhibited yeast cells. This control population was then used to set an empirical False Discovery Rate at which an enrichment ratio is not expected to occur by chance in a population of a given size.

The most sophisticated approach for analysis came from the Bloom group, who sought to quantitatively estimate binding dissociation constants for S RBD mutants. Their approach involved sorting using many different labeling concentrations of ACE2 or antibody and using a maximum likelihood estimation approach to infer dissociation constants [68]. This protocol is very exhaustive, with precision coming at the expense of throughput. Thus, this is a suitable protocol to analyze a few antibodies in depth.

In summary, these three groups' contributions show how different experimental observables result from different experimental strategies.

2.8   Conclusions and perspective

Hundreds of antibodies and nanobodies have had their epitopes mapped on SARS-CoV-2 S with a handful also having their escape mutants determined. This accumulated knowledge has contributed to the mitigation of COVID-19 through the development of monoclonal antibody therapies and novel vaccines. All techniques surveyed were quite useful for different facets of epitope mapping. Electron microscopy and X-ray crystallography were essential in the early days by defining the structural basis of many common neutralizing and non-neutralizing epitopes on S. Linear peptide arrays showed the diversity of the antibody response against S and identified several common immunodominant epitopes. Deep mutational scanning was essential for understanding the impact of individual mutations on both S RBD/ACE2 and anti-S antibody/S recognition. These mutational constraints on binding led to predictive understanding of the recognition landscape for currently circulating VoCs.

Our mini-review described at length different conformational epitope mapping methods by deep mutational scanning as no in-depth review for this methodology exists. We are especially excited about the ability to delineate the sequence constraints on binding by both ACE2 and nAbs, as these constraints dictate the boundaries of the emerging arms race between future mutations on SARS-CoV-2 VoC and the ability of the humoral response in the vaccinated and naturally infected population to respond. It remains to be seen whether deep mutational scanning can inform the next generation of design of monoclonal antibody therapies and vaccine candidates.

2.9   Acknowledgements

2.10   Author Contributions

Wrote paper: IMFU, TAW.

Chapter 3

One-shot identification of SARS-CoV-2 S RBD escape mutants using yeast screening

The potential emergence of SARS-CoV-2 Spike (S) escape mutants is a threat to reduce the efficacy of existing vaccines and neutralizing antibody (nAb) therapies. An understanding of the antibody/S escape mutation landscape is urgently needed to preemptively address this threat. Here we describe a rapid method to identify escape mutants for nAbs targeting the S receptor binding site. We identified escape mutants for five nAbs, including three from the public germline class VH3-53 elicited by natural COVID-19 infection. Escape mutations predominantly mapped to the periphery of the ACE2 recognition site on the RBD with K417, D420, Y421, F486, and Q493 as notable hotspots. We provide libraries, methods, and software as an openly available community resource to accelerate new therapeutic strategies against SARS-CoV-2.
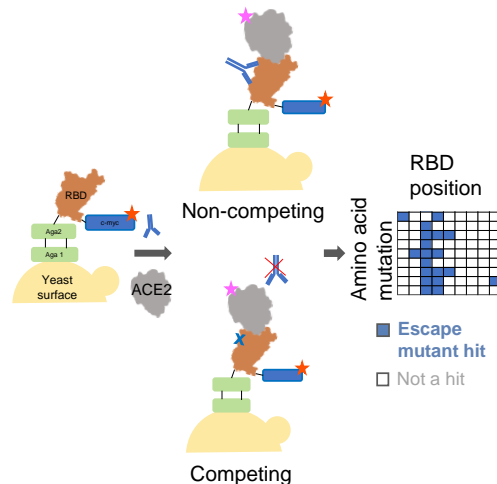
3.1   Introduction

The type I viral fusion protein Spike (S) is a major antigenic determinant of SARS-CoV-2 and is the antigen used in all approved COVID-19 vaccines [75, 76, 77]. Recently, the B.1.1.7 (N501Y; Alpha), B.1.351 (E484K, N501Y, K417N; Beta), B.1.427 (L452R; Epsilon), B.1.617 (L452R, E484Q; Delta) and C.37 (L452Q, F490S ;Lambda) viral lineages have emerged (mutations listed are for S receptor binding domain (RBD) only). Among other mutations on Spike, all lineages encode single nucleotide substitutions in the S RBD near the recognition site for its cellular target angiotensin-converting enzyme 2 (ACE2) [77, 78, 79].

Dozens of studies have reported the structural, epitopic, and functional landscape of non-neutralizing monoclonal antibodies and nAbs targeting trimeric S [2, 4, 28]. A prophetic under-standing of the mutations on S that could evade antibody recognition would enable development of better vaccine boosters and monoclonal antibody therapies. In particular, FDA-approved mon-oclonal antibody therapies targeting the S RBD developed by Regeneron [30] and Lilly [58] have shown significantly decreased effectiveness with Beta and Gamma variants [80, 31]. Thus, we sought to develop an S RBD yeast surface display (YSD) platform (**Figure C.2**) [35], as we hypothesized that broad identification of SARS-CoV-2 S escape mutants could be found by integrating high throughput screening platforms with deep sequencing. While a similar platform uses the loss of nAb binding to identify escape mutants [31, 65], we rationalized that a functional screening assay that directly measures the ability of a nAb to compete with ACE2 for S RBD binding, would be a comparatively strong predictor of RBD escapability, as it accounts for mutations in RBD that would disrupt S binding to ACE2.

3.2   Results

We have developed an aglycosylated S-RBD YSD platform (S RBD(333-537)-N343Q) from the original Wuhan-Hu-1 strain [2] that can bind specifically to ACE2 **(Figure 3.1A)**.This S RBD construct has its one native N-linked glycan removed (N343Q) as the heavy N-linked mannosylation

endemic of S. cerevisiae could hamper anti-S RBD mAb recognition [72]. With it, we characterized the epitope of the monoclonal antibody 910-30 and confirmed that it binds a glycan-indepenent epitope. By performing a competition binding experiment against ACE2, we determined that it directly competes against ACE2 (**Figure C.1**) for binding, like ofther IGHV3-53/3-66 class members. Cell surface titrations of CR3022 IgG and nAb HKU-910-30 IgG yielded apparent dissociation constants comparable to reported in vitro results [2, 28] (**Figure C.3**). We next tested a panel of eleven additional anti-S RBD mAbs for binding to aglycosylated RBD. These mAbs were isolated from convalescent donors infected in late 2019/early 2020 and thus are representative of anti-S mAbs raised during natural infection [4]. Ten of the eleven mAbs recognized aglycosylated S RBD (**Figure 3.1B**). The one panel member that did not bind, CC6.33, selectively recognizes the S309 epitope on the RBD containing the N-linked glycan at position 343 [36].

Next, we evaluated the ability of the mAb panel to competitively inhibit ACE2 binding to aglycosylated S RBD in an assay conceptually similar to the one previously described by Tan et al. [81]. Yeast displaying aglycosylated S RBD was first labeled with a saturating concentration of a given mAb and then co-incubated with biotinylated ACE2. Six mAbs completely ablated ACE2 binding, one mAb partially inhibited ACE2, and the remaining four did not prevent ACE2 binding (**Figure 3.1C**). A direct correlation was observed between the previously determined neutralization potency of the antibody [2, 28] and the fluorescence signal increase in the competition assay (**Figure 3.1C**). We conclude from these experiments that, excluding the S309 epitope, the aglycosylated S RBD platform faithfully recapitulates binding interactions of nAbs with S RBD [4].

Our strategy for identifying potential S RBD escape mutants was as follows. First, we constructed a saturation mutagenesis library of aglycosylated S RBD containing all possible single missense and nonsense mutations for the 119 surface exposed positions of the RBD (96% coverage of the 2,380 possible library members; **Table D.1** contains library coverage statistics) [82]. For each codon, mutations were encoded using oligonucleotides containing a degenerate NNK sequence. This degenerate sequence encodes all 20 amino acids plus a stop codon, which is useful as an internal
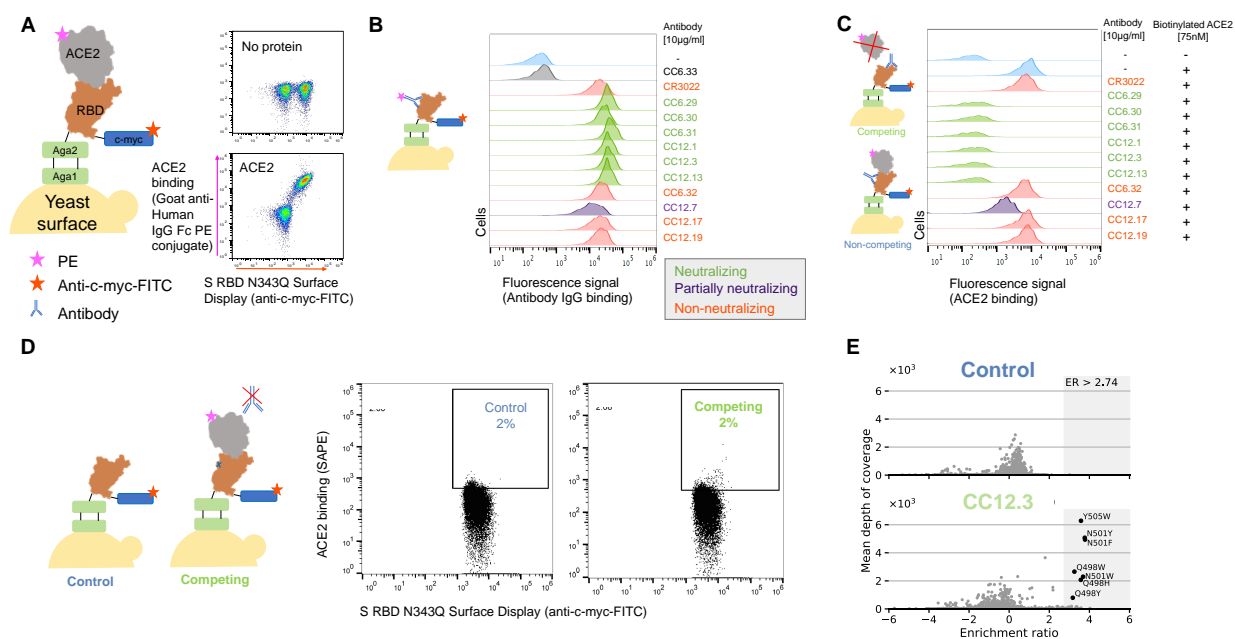
Figure 3.1: **Identification of SARS-CoV-2 S RBD escape mutants using yeast screening. A.** Cartoon of the yeast display construct S-RBD(333-537)-N343Q. Cytograms show specific binding in the presence, but not absence, of ACE2-Fc. **B.** Binding profiles of aglycosylated S RBD labeled with 10 μg/mL of indicated mAb. Antibodies are color coded according to neutralization potency [4]. **C.** Competitive binding between IgG and ACE2 was performed by labeled yeast displaying aglycosylated S RBD with 10 μg/mL of indicated mAbs followed by labeling with biotinylated ACE2. **D.** Single site saturation mutagenesis S RBD libraries were sorted by FACS using a competition experiment. The top cytogram shows the cell population collected for the control population without ACE2 labeling, while the bottom cytogram shows the cell population enriched in mutations able to bind ACE2 in the presence of a competing IgG. The specific cytogram shown is for nAb CC12.3 using the S RBD library corresponding to mutations at positions 437-537. **E.** Per-mutation enrichment ratio distributions as a function of average depth of coverage control (top) and CC12.3 nAb competing experiment (bottom). See also **Figures C.3, C.4, C.5, C.6**.

negative control for the assay. We labeled yeast displaying these RBD variants with a saturating concentration of nAb and then co-incubated with a saturating concentration of biotinylated ACE2. We then used fluorescence activated cell sorting (FACS) to screen for mutants that could bind ACE2, indicating that the RBD mutation allows for evasion of the nAb while not disrupting the ACE2 interaction critical for cell entry **(Figure 3.1D)**, **Figure C.4, C.5**). Importantly, a control with no ACE2 labeling was sorted to set an empirical false discovery rate (FDR) for putative escape mutant hits (**(Figure 3.1E)**, **Figure C.5**). Plasmid DNA from sorted cells were prepped and deep sequenced. We determined the enrichment ratio (ER) – the base-2 logarithm of the ratio of the frequency mutant in the sorted population to its frequency in the reference population – and then used the control population to set the FDR **(Figure 3.1E)**, **Fig C.6**). We screened five different nAbs identified earlier as having completely ablated ACE2 binding (CC6.29, CC6.31, CC12.1, CC12.3, CC12.13). In all, we identified a total of 97 S RBD mutants that can escape recognition by at least one nAb (**Table D.2**).

For all five nAbs, the putative escape mutant hits were localized to specific locations within the S RBD primary sequence **(Figure 3.2A**, **Figure C.7, C.8, C.9, C.10, C.11)**. CC12.1 and CC12.3 belong to the public germline class VH3-53 [2, 38, 28] and are representative of the subset of VH3-53 public antibodies with relatively short CDRH3 regions [83]. Strikingly, these two nAbs share over 90% of the same RBD escape mutants **(Figure 3.2B)**, even though the light chain differs between the nAbs. Structural complexes of antibodies CC12.1 and CC12.3 were previously solved in complex with S RBD [37], affording a structural explanation for individual escape mutants. Escape mutants for both of the VH3-53 nAbs CC12.1 and CC12.3 clustered at the same location on the S RBD mainly on the periphery of the ACE2 binding site **(Figure 3.2C**; **Figure C.12)**. To confirm that the mutations did not have a large effect on equilibrium binding to ACE2, we determined the dissociation constant of eight single point mutants using yeast surface display titrations. Binding affinities of each mutant were comparable to the S RBD N343Q dissociation constants and in agreement with a previous deep mutational scanning study (10) **(Figure 3.2D**; Data S3). Thus,

mutations identified that escape antibody recognition in this assay can still bind ACE2.

Having identified a number of putative escape mutants from the mutagenesis library screening, we sought to determine how this functional screening correlated with the more conventional pseudovirus neutralization assay. A panel of MLV-based SARS-CoV-2 pseudoviruses were generated that contained single mutations predicted by the mutagenesis scanning to allow escape from one of the antibodies screened, as well as several irrelevant control mutations. Antibodies CC12.1, CC12.3 and CC6.29 were screened against the original SARS-CoV-2 pseudovirus as well as this panel of mutant pseudoviruses in duplicate **(Figure 3.2E)**, and the resulting IC50 values were compared to calculate the effect on antibody neutralization potency **(Figure 3.2F**; **Figure C.13)**. Consistent with the RBD mutagenesis library and structural analysis, CC6.29 failed to neutralize the F486I, E484K, and T478R variants. Additionally, K417N, K417T, and D420K hotspot mutants completely escaped neutralization for both CC12.1 and CC12.3. The only instance we tested where the mutagenesis scanning data differed from the pseudovirus results was at N501Y that was predicted to confer escape from CC12.1 and CC12.3 but had no effect on the in vitro neutralization potency. Although it is unclear why this discrepancy occurred, we note that N501Y significantly increases the affinity of the RBD for ACE2, which could result in ACE2 out-competing bound nAbs.

Finally, we performed biological replicates where the mutagenesis library corresponding to S RBD positions 437-537 was separately transformed into yeast and screened against nAbs CC6.29, CC12.1, and CC12.3. While the enrichment ratios were lower than in the initial experiment, nearly the same set of escape mutants was identified for CC6.29, and escape mutants originally identified for all nAbs had significantly higher ERs than other variants in the replicate (p-value range 4.2e-4 to 1.9e-11, one-sided Welch's t-test) **(Figure C.14)**.

Selected per-position heatmaps, and structural mapping of S RBD escape mutants, are shown in **Figure 3.3** for all five nAbs. Closer examination of these datasets reveals key features of the RBD escape mutant response. CC12.1 and CC12.3 nAbs share over 90% of the same RBD escape
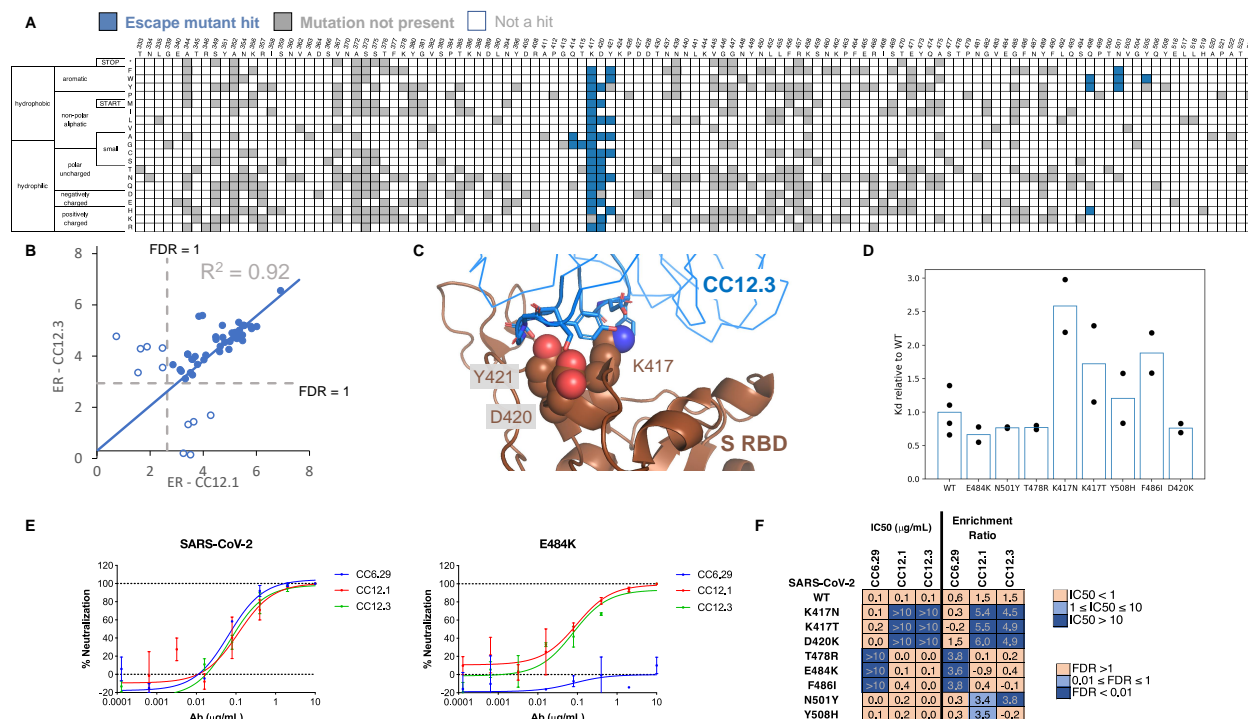
Figure 3.2: **Validation of escape mutants using yeast screening and pseudo neutralization assays.** **A.** Heatmap showing predicted S RBD escape mutants for CC12.3 in blue. White cells are mutations with a p value for an FDR ¿ 1, while grey cells are mutations not present in the mutational library. S RBD positions directly involved in binding ACE2 are colored gray. **B.** Comparison of enrichment ratios (ER) for individual hits for CC12.3 vs. CC12.1. Closed circles represent escape mutant hits for both nAbs, whereas open circlers are escape mutant hits for only one nAb. **C.** Solved structure of nAB CC12.3 in complex with S RBD (PDB ID 7KN6). **D.** Dissociation constants of single point mutants relative to S RBD N343Q ("WT") determined by yeast surface display titrations. Circles show the relative value for each biological replicate and the bars represent the mean (n=2 for each mutant; n=4 for WT). **E.** PSV neutralization curves for CC12.1, CC12.3 and CC6.29 on SARS-CoV-2 (top) and SARS-CoV-2 E484K (bottom). **F.** PSV IC50 analysis for CC12.1, CC12.3 and CC6.29 on different identified mutations. See also **Figure C.7, C.8, C.9, C.10, C.11, C.12, C.13, C.14**, and **Table D.2**.

mutants (**Figure 3.2B**), including notable hotspot mutations occurring at K417, D420, Y421, and Q498 (**Figure 3.3A**). Interestingly, multiple aromatic substitutions at Q498 escape recognition for CC12.1 and CC12.3 even though the antibodies have different light chains and recognition motifs for that position. Introduction of an aromatic residue at Q498 introduces substantial van der Waals clashes that are likely unresolved without antibody loop movement. The other VH3-53 nAb tested, CC12.13, has a 15 amino acid length CDRH3 that likely has a distinct binding mode than that for CC12.1 and CC12.3 [83]. Consistent with this, the CC12.13 escape mutants identified are mostly different from those for CC12.1/CC12.3 (**Figure C.11**).

Another nAb screened, CC6.29, has a completely different escape mutant profile compared with CC12.1/CC12.3. The 15 potential RBD escape mutants for CC6.29 center around the structural 'knob' of positions A475, S477, T478, E484, and F486 (**Figure 3.3B**). E484K shared by the B.1.351 and B.1.526 lineages is identified as an escape mutant for this nAb, but the structurally adjacent S477N mutation newly identified in the B.1.526 lineage does not escape CC6.29 neutralization. Intriguingly, S477P is identified as an escape mutant for this nAb. F486 is a mutational hotspot even though that position is involved in the recognition of ACE2. This is consistent with a previous mutational scan of S RBD showing that mutation of F486 does not significantly impact ACE2 binding affinity [65]. CC6.31 escape mutants partially overlap with CC6.29 but implicate a different set of mutants (**Figure 3.3B**). Multiple mutations at Q493 escape CC6.31, including Q493 substitutions to the aromatic amino acids F/W.

In total, the five nAbs map a partially overlapping surface with the ACE2 binding site that is primed for antibody escape. In comparison with the binding footprint of ACE2 (**Figure 3.3C**), the escape mutants almost completely map to the outer binding shell and periphery of the interaction surface, akin to an O ring circumscribing the receptor binding site (RBS). Out of the identified escape mutants, residues K417, F486, Q493, N501, and Y505 are located on the ACE2 footprint (**Figure 3.3C**). While mutations on K417 and F486 do not significantly change the RBD affinity to ACE2, mutations on N501 can increase or decrease affinity depending on the substitution. The
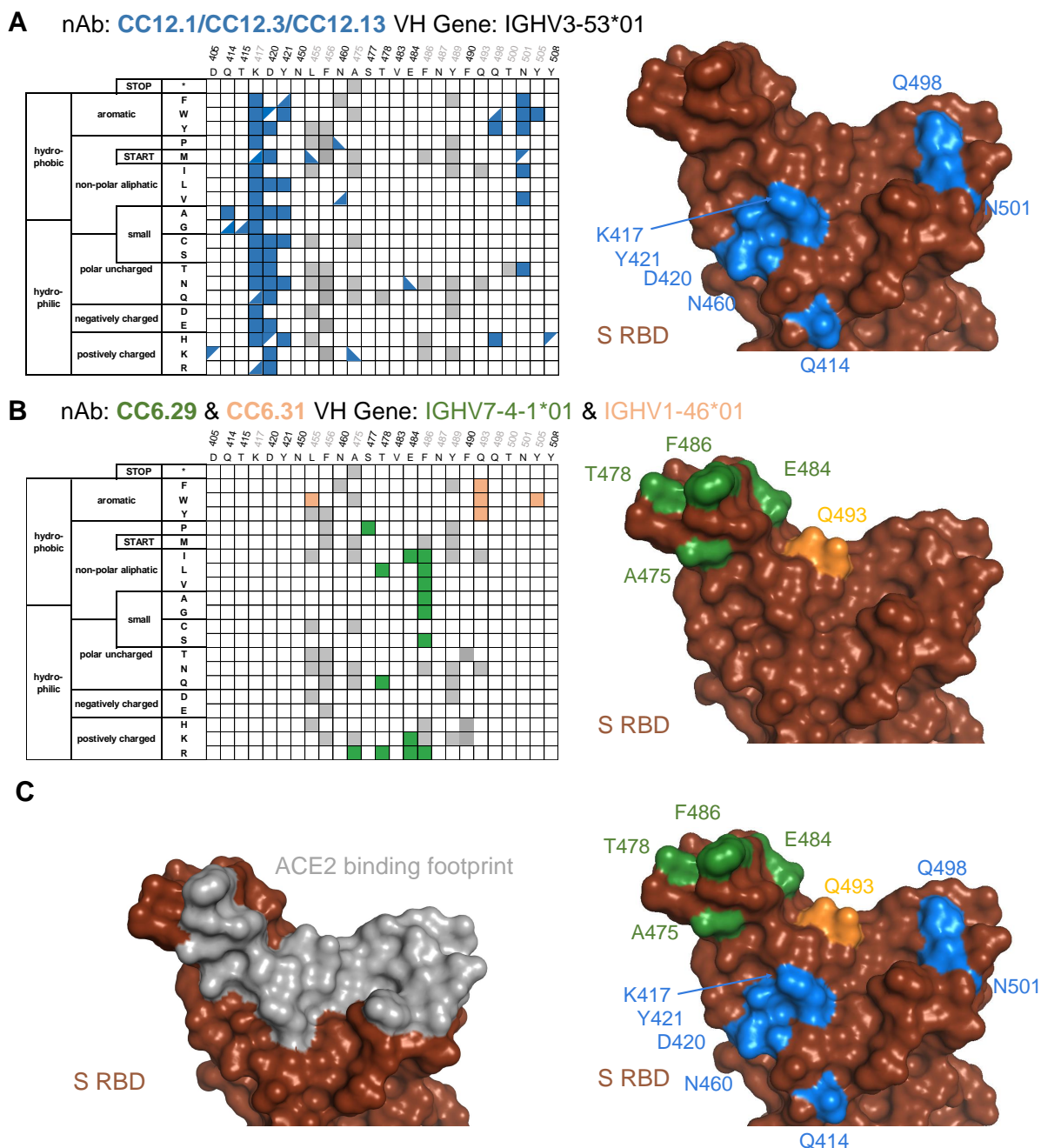
Figure 3.3: **Sequence determinants and structural basis of S RBD escape mutants. A-C.** (left) Limited per-position heatmap and (right) mutations mapped onto the S RBD-ACE2 structural complex (PDB ID: 6M0J). For clarity, only positions with two or more escape mutations are shown with surface colored. Panels are for **A.** nAbs CC12.1, CC12.3, and CC12.13. Boxes indicate escape mutants for two or more nAbs, while triangles indicate an escape mutant identified for just one nAb (top left: CC12.1, bottom right: CC12.3, bottom left: CC12.13); **B.** CC6.29 and CC6.31 (orange); **C.** overlay of escape mutants from all nAbs onto the S RBD-ACE2 structural complex.

Y505W mutant shared by CC6.31, CC12.1, and CC12.3 also increases ACE2 affinity [65].

We were puzzled by the fact that the mutations at D420 were so deleterious to the neutralization potency of the VH3-53 nAbs given that this residue is on the outer periphery of the binding epitope. Consequently, we performed 100 ns aqueous molecular dynamics (MD) simulations of CC12.1 and CC12.3 in complex with wildtype S RBD. We also simulated S RBD incorporated with the D420E, D420K, or the Y421N mutation (see Method details for details). In the control simulation with CC12.1, D420 on the RBD and CDRH2 S56 on CC12.1 form persistent hydrogen bonds, and Y421 on the RBD is tightly bound within a pocket of CC12.1 residues (**Figure 3.4A**). With the D420E mutation, the increased length of E420 disrupts its ability to hydrogen bond with S56, requiring it to adopt a bent conformation (**Figure 3.4B**). This forces Y421 out of the antibody pocket, causing increased fluctuations in neighboring RBD loops which persist throughout the entire 100 ns production simulation (**Figure C.15a, b**). With the D420K mutation, hydrogen bonding with S56 is completely disrupted. With the Y421N mutation, N421 is too short to interact with the antibody pocket (**Figure C.15c**). Similar escape mechanisms are observed for CC12.3 with all three RBD mutations, including increased fluctuations at one of the same key sites (K458) on the RBD in response to the D420E mutation (**Figure C.15d, e**).

There have been a number of recent approaches to identify specific S escape mutants (summarized in **Table D.3**) [30, 31, 84, 85, 86]. A survey of the existing escape mutant literature, along with escape mutants identified in the present work, allows us to identify the absolute and near-absolute escape resistant ACE2 RBS residues in the context of the original lineage (**Figure 3.4C**). One resistant patch is found around F456/Y473/N487/Y489 while other residues are discontinuous patches on the remainder of the RBS. We note that many of these same resistant residues are identical to those from SARS-CoV (Y449, N487, Y489, G496, T500 and G502). The lack of a contiguous surface at the RBS that is conserved makes it highly unlikely that one could identify a naïve nAb targeting the RBS that is completely resistant to escape.

A

CC12.1

RBD

S56

Y421   D420

B

D420E

C Escape Mutants at Receptor Binding Site

Y489   G496

G447   T500

Y449

N487

Y473

F456

G502

Y453   Y495   F497

■ >1 mutation ■ One mutation ■ No mutations

D

Position	Variant	nAb CC12.1	CC12.3	CC12.13	CC6.29	CC6.31	Lineage	Observed	ACE2 binding
333	T333K	■							
358	I358F	■	■						□
415	T415R								
417	K417E	■						■	□
	K417I	■							
	K417M	■							
	K417N	■						■	□
	K417Q								
	K417R							■	□
	K417T								
420	D420A	■							
	D420E	■							
	D420H	■							
	D420N	■							
	D420Y	■							
421	Y421C	■							
	Y421F	■							
	Y421H	■							
	Y421N	■							
455	L455M					■			
	L455W					■			
460	N460H		■						
	N460I		■						
	N460K	■	■						□
	N460S		■						
	N460T		■						
473	Y473F		■						
475	A475E		■						□
477	S477P			■					□
478	T478P				■			■	
	T478K				■				
	T478R				■				
484	E484D				■			■	□
	E484K				■		■	■	□
	E484V				■				
485	G485V				■				□
486	F486C				■				
	F486I				■				
	F486L				■			■	
	F486S				■			■	
	F486V				■			■	□
487	N487D								□
490	F490C					■			
	F490S					■		■	
493	Q493H					■			
	Q493R					■		■	
494	S494F					■			
498	Q498H	■							■
501	N501I	■		■					□
	N501T			■					□
	N501Y			■			■		■
503	V503E								
508	Y508H	■						■	□
527	P527L								□

■ Identified in this study   ■ Higher ACE2 affinity
■ Present in current strain   ■ Neutral
■ Observed in literature   ■ Lower ACE2 affinity

Figure 3.4: **Mechanistic, structural, and sequence analysis of SARS-CoV-2 escape mutants. A-B:** Snapshots from MD trajectories showing **A.** key interactions in the control simulation of S RBD in complex with CC12.1, and **B.** mechanism of escape of S RBD from CC12.1 due to the D420E mutation. Images were rendered with Visual Molecular Dynamics (VMD (Humphrey et al., 1996)), and black dotted lines indicate persistent hydrogen bonds. **C.** S RBS positions are colored by the number of escape mutants identified to date. RBS residues involving the S RBD-ACE2 structural complex (PDB ID: 6M0J) are colored by number of escape mutants identified to date. **D.** Summary of 1-nt escape mutants identified in the present study. Lineage column indicates presence of the given mutation amongst currently circulating SARS-CoV-2 strains, while the observed column refers to an escape mutant previously identified in literature [30, 31, 84, 85, 86]. ACE2 binding indicates affinity to ACE2 based on the measurements by Starr et al. [65]. See also **Figure C.15** and **Table D.3**.

A major near-term concern with public health implications is identification of the set of single nucleotide polymorphisms that encode for escape mutants on the S RBD. A summary of 1-nucleotide (nt) escape mutants identified in the present work is shown in **Figure 3.4A**. To our knowledge, 40/54 (74%) of 1-nt escape mutants identified from this nAb panel have not previously been identified, including hotspot positions D420 and Y421 that escape recognition by the abundant VH3-53 nAbs. Other notable residues identified here include S477, Q498, and Y501, as these positions lie directly on the RBS and all have been shown to slightly increase binding affinity to ACE2 [65]. Mutants K417N, E484K, and N501Y in currently circulating lineages escape some, but not all, of the nAbs on the panel.

The current study has been performed in the context of the original Wuhan-Hu-1 strain. Nonetheless, new variants are emerging and further research should be conducted to gain insight on the escape mutants in the presence of multiple mutations. To that end, we have constructed new libraries containing a constant mutation to E484K and N501Y present in the Alpha, Beta and Iota variants of concern (88.7% and 91.8% library coverage respectively; **Table D.1** contains library coverage statistics).

3.3   Discussion

We have developed a yeast platform that allows for the rapid identification of SARS-CoV-2 S RBD escape mutants for a given nAb. While other platforms to identify escape mutants have recently been described, key advantages of the approach presented here include: (i.) screening by competitive binding against ACE2 which more precisely mimics how actual viral infection can still persist despite antibody binding; (ii.) a robust and rigorous hit identification algorithm; (iii.) a safe working environment, as it does not use live virus; and (iv.) a relatively fast identification, as the RBD library can be screened against a given nAb and analyzed in under a week.

3.4   Limitations of study

There also exist drawbacks. First, the present method is limited to mapping escape mutants for anti-S-RBD nAbs that directly compete with ACE2 for binding. Many nAbs neutralize by targeting S epitopes across protomers [50] or on the N-terminal domain [41], and a robust platform for S ectodomain display would enable more comprehensive studies. We attempted to develop a yeast surface display platform for the full S ectodomain but were unsuccessful: we screened media composition, expression temperature, protein orientation (**Figure C.16**), and mutations (1,909 mutants screened with only two potential hits) (Data S3, **Table D.1**, **Figure C.17**). Second, the presented assay measures the ability of a given mutant to escape nAb blockade of ACE2. While from all available data the assay appears to correlate well in the context of pseudo-virus, each mutation is pleiotropic with unknown fitness effects beyond escape for a given nAb; the true RBS escape mutants that do not appreciably impede viral fitness will be a subset of the mutations identified here.

Still, using this method we were able to identify specific failure mechanisms for five different nAbs. This tool can be easily adapted and contribute to developing the next generation of broadly neutralizing antibodies against SARS-CoV-2, as well as suggest mutations to include for the next generation of vaccines. The two major prospective applications for this tool then are for monoclonal antibody therapy and universal vaccine design against SARS-CoV-2 (i.e., generating vaccine-elicited antibodies that are resistant to viral escape). The rationale for using this tool in the context of monoclonal antibody therapy is arguably stronger, as FDA-approved therapies like bamlanivimab [58], among others, are not as effective against currently circulating variants. The antibodies used here are from convalescent patients and represent antibodies raised during natural infection. While some FDA-approved antibodies were not derived from convalescent patients, in principle any nAb that directly competes with ACE2 binding should be amenable to this technique. We have developed mutagenesis libraries in three different RBD backgrounds, and new libraries could be developed to match genotypes for future variants of concern.

On the other hand, it remains to be seen whether this yeast platform could be used in the context of universal vaccine design, as individual nAbs, or combinations thereof, are often not representative of bulk sera. Thus, it would be interesting to see whether our yeast platform presented here is robust enough to identify escape mutants from bulk sera from convalescent or vaccinated individuals.

3.5 STAR Methods

3.5.1 Resource availability

3.5.1.1 Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Timothy A. Whitehead (timothy.whitehead@colorado.edu).

3.5.1.2 Materials availability

Pooled libraries and plasmids from this study will be available at Addgene: `https://www.addgene.org/Timothy_Whitehead/`

3.5.2 Experimental model and subject details

Appendix A provides a detailed step by step protocol for this method.

3.5.2.1 Cell lines

Saccharomyces cerevisiae strain EBY100 (ATCC MYA-4941TM) was cultured at 30°C for cell growth, and at 22°C for cell induction in flasks while shaking at 300rpm. Cells were incubated in 6.7g/L Difco yeast nitrogen base, 5g/L Bacto casamino acids, 5.4g/L Na_2HPO_4, and 8.56g/L $NaH_2PO_4 \cdot H_2O$ and 20g/L carbon source (dextrose for cell growth and galactose for cell induction). HEK293T (ATCC CRL-3219) were cultured in DMEM (Corning 15-013-CV) with 10%

heat-inactivated FBS, 2 mM L-glutamine and 1X PenStrep at 37°C in a humidified 5% CO2 incuba-

tor. Vero-E6 cells (ATCC CRL-1586) were plated in a T225 flask with complete DMEM (Corning

15-013-CV) containing 10% FBS, 1X PenStrep, 2 mM L- Glutamine (Corning 25-005-CL) overnight

at 37°C and 5% CO2. HeLa-ACE2 cells were seeded in 12 μL complete DMEM at a density of

$2x10^3$ cells per well.

3.5.2.2   Bacterial strains

Escherichia coli strain XL1-Blue and Mach1TM were incubated in LB media at 37°C and

300rpm in culture tubs.

3.5.3   Method details

3.5.3.1   Plasmid constructs

All plasmids and all primers used for this work are listed in Key resources table. All plas-

mids were verified by Sanger sequencing. Yeast display constructs for SARS-CoV-2 spike protein

ectodomain (GenBank MN908947 with a GSAS substitution at the furin cleavage site (682-685)

and proline substitutions at positions 986 and 987 ([35]), and a C-terminal T4 fibritin trimerization

domain), as shown in **Figure C.2**, were constructed as follows. Spike was codon optimized for

Saccharomyces cerevisiae with Benchling software using default options, split into three gene blocks

(hereafter labeled A, B, and C) each encoded with BsaI restriction sites with overhangs ([87]), syn-

thesized as gBlocks (IDT), and cloned into pUC19 (Addgene: #50005) using SalI/KpnI restriction

sites. This yielded the spike fragment entry plasmids pUC19-S-ecto-B and pUC19-S-ecto-C-Nterm.

To construct pUC19-S-ecto-A-Nterm-KanR (the spike fragment destination plasmid), PCR was

used to amplify both the kanamycin resistance gene from pETconNK (Addgene: #81169) with

primers MBK-175 and MBK-176, and the pUC19-S-ecto-A-Nterm plasmid with primers MBK-177

and MBK-178. NEBuilder HiFi DNA Assembly protocol (NEB) was used to insert the kanamycin

resistance gene into the plasmid. pUC19-S-ecto-Nterm was constructed by Golden Gate cloning

([88]) using pUC19-S-ecto-A-Nterm-KanR, pUC19-S-ecto-B, and pUC19-S-ecto-C-Nterm.

To construct pJS698 (N-terminal fusion Spike ectodomain YSD backbone), pETconNK-Nterm-Aga2p was first constructed by inserted a gene block with a multiple cloning site between the AGA2 signal peptide and the remainder of the AGA2 coding sequence following standard restriction enzyme cloning practices. pETconNK-Nterm-Aga2p was amplified with primers PJS-P2194 and PJS-P2195 using KAPA HiFi HotStart Readymix (Kapa Biosystems). The reaction was fractionated by agarose gel electrophoresis and the 6062 bp band excised and purified using a Monarch DNA Gel Extraction kit (NEB). The fragment (40 ng) was circularized using the Q5® Site-Directed Mutagenesis Kit (NEB) in a 10 μl reaction and transformed into E. coli Mach1 chemically competent cells (Invitrogen).

To construct pJS697 (C-terminal fusion RBD YSD backbone), pETconNK (Addgene: #81169) was amplified with primers PJS-P2192 and PJS-P2193 using KAPA HiFi HotStart Readymix (Kapa Biosystems). The reaction was fractionated by agarose gel electrophoresis and the 6084 bp band excised and purifed using a Monarch DNA Gel Extraction kit (NEB). The fragment (40 ng) was circularized using the Q5® Site-Directed Mutagenesis Kit (NEB) in a 10 μl reaction and transformed into E. coli Mach1 chemically competent cells (Invitrogen).

pJS699 (Wuhan-Hu-1 S-RBD(333-537)-N343Q for fusion to the C-terminus of AGA2) was previously described ([2]). S RBD Single point mutants were introduced following the Kapa HiFi Hotstart Readymix protocol (Cat# 7958927001) with the following protocol:

98°C	3min	
98°C	20s	
Melting temperature of each primer	15s	25 cycles
72°C	2:40min	
72°C	3:40min	
4°C	Hold	

Primers used in site directed mutagenesis are given in Key resources table. Amplicons were fractionated by agarose gel electrophoresis and purified using a Monarch DNA Gel Extraction Kit (NEB). Further ligation of the purified amplicons was performed using T4 ligase and PEG. Finally, the plasmids were transformed into E. coli Mach1 cells and incubated overnight. On the following day the DNA was extracted using an NEB Miniprep Kit. pIFU001 - pIFU008 (Key resources table) contain the single mutants E484K, N501Y, T478R, K417N. K417T, Y508H, F486I and D420K respectively.

3.5.3.2   Recombinant protein production, purification, and preparation

ACE2-Fc was produced and purified following Walls et al. 2020 ([32]). CR3022 ([89]) was expressed by transient transfection in Expi293F cells and purified by protein A affinity chromatography and SEC using a Superdex 200 10/300 GL. Specificity was verified by measuring binding to SARS-CoV-2 RBD and irrelevant antigen. The anti-SARS-CoV-2 RBD antibody panel used (CC6.29, CC6.31, CC6.32, CC6.33, CC12.1, CC12.3, CC12.7, CC12.13, CC12.17, CC12.19) was a kind gift from Dennis Burton's lab at Scripps and were produced and purified according to Rogers et al. ([4]).

All proteins that were chemically biotinylated were prepared at a 20:1 molar ratio of biotin to protein using EZ-Link NHS-Biotin (Thermo Scientific) according to the manufacturer's protocol. All proteins were stored at 4°C in phosphate buffered saline (8 g/L NaCl, 0.2 g/L KCl, 1.44g/L

Na_2HPO_4, 0.24g/L KH_2PO_4) pH 7.4.

3.5.3.3   Preparation of Mutagenic Libraries

All 119 surface exposed positions on S RBD (positions 333-537) were mutated to every other amino acid plus stop codon using NNK primers (Key resources table) using comprehensive nicking mutagenesis exactly as described (**Wrenbeck et al., 2016**). For compatibility with Illumina sequencing, two tiles were made: tile 1 encompassed positions 333-436, while tile 2 encompassed positions 437-527 containing the critical receptor binding site. Serial dilutions were plated to calculate the transformation efficiency (**Table D.1**).

To create the display construct of S-RBD(333-537)-N343Q fused to the C-terminus of Aga2p, pJS697 was digested with BsaI-HFv2 (NEB) and purified using a Monarch PCR & DNA Cleanup Kit (NEB). Each mutated pJS699 library was digested with NotI-HF (NEB), the reaction fractionated by agarose gel electrophoresis, and the band corresponding to S-RBD (0.83kb) excised and purified using a Monarch DNA Gel Extraction Kit (NEB). Yeast transformation was performed exactly as described ([90]). For each library, the two fragments were co-transformed (in a 3:1 molar ratio of S-RBD to backbone) into chemically competent S. cerevisiae EBY100 ([91]). Serial dilutions were plated on SDCAA and incubated 3 days to calculate the efficiency of the transformation (**Table D.1**). Biological replicates were made on a different day by co-transforming each tile into EBY100 exactly as described. Yeast stocks for each transformation were stored in yeast storage buffer (20 w/v % glycerol, 200 mM NaCl, 20 mM HEPES pH 7.5) at -80°C.

Mutagenic libraries for the N-terminal spike orientation were constructed following oligo pool mutagenesis exactly as described ([92, 82])) using pUC19-S-ecto-A-Nterm-KanR, pUC19-S-ecto-B, and pUC19-S-ecto-C-Nterm as templates. For the oligo pool we computationally selected 1,909 mutations hypothesized to either destabilize the 'down' conformation, stabilize the 'up' conformation, or both (Data S3). The majority of these mutations targeted S1 (94%, 1793/1909) at

the NTD, RBD, SD1, and SD2 domains, with the remainder mapping to the boundary between the HR1 and CH domains on S2. After mutagenesis, the mutational libraries were digested with BsaI-HFv2, fractionated by agarose gel electrophoresis, and gel excised and purified with Monarch Gel Extraction kit (NEB). 40 fmol of pUC19-S-ecto-A-NSM-Nterm-KanR, pUC19-S-ecto-B-NSM, and pUC19-S-ecto-C-NSM-Nterm were ligated together with T4 DNA Ligase (NEB), cleaned up and concentrated each to a final volume of 6μl with Monarch PCR & DNA Cleanup kit (NEB), and transformed into chemically competent E.coli Mach1 cells (Invitrogen cat. #C862003). The resulting library had on average 3 mutations per spike protein per plasmid. Library statistics were determined post sequencing.

To construct the surface display library in yeast, the spike plasmid library was digested with NotI-HF (NEB) and the S coding region was gel purified. The YSD vector pJS698 was digested with BsaI-HFv2 and column purified. 1.3 μg of insert (S coding region) and 1.7 μg of vector were electroporated into 400 μl EBY100 using the method of Benatuil et al. ([93]) as written, except that electroporation was performed at 2 kV rather than 2.5 kV. Serial dilutions were plated on SDCAA Agar to calculate the complexity of the library. After recovery, the cells were transferred to 50 ml SDCAA (20g/L dextrose, 6.7g/L Difco yeast nitrogen base, 5g/L Bacto casamino acids, 5.4g/L Na_2HPO_4, and 8.56g/L $NaH_2PO_4 \cdot H_2O$) and grown at 30°C for two days to saturation. The cultures were passaged twice in medium M37D (diluted to $OD_600 = 0.05$ in 120 ml, then to $OD_{600} = 0.4$ in 50 ml) and stocks prepared at $OD_{600} = 1$ as in Whitehead et al. ([94]). The final composition of M37 is 20 g L^{-1} dextrose or galactose (for M37D, M37G respectively), 5 g L^{-1} casamino acids, 6.7 g L^{-1} yeast nitrogen base with ammonium sulfate, 50 mM citric acid, 50 mM phosphoric acid, 80 mM MES acid, neutralized with 90% sodium hydroxide / 10% potassium hydroxide to pH 7. Both media should be prepared by dissolving all reagents except yeast nitrogen base into MilliQ water, adjusting the pH to 7.0 with freshly prepared sodium hydroxide / potassium hydroxide mixture, and adjusting the volume to 9/10th of the final desired volume. Pass the solution through a 0.22 μm filter, both for sterility and to remove particulates that would

nucleate struvite. Finish the media by addition of 1/10th volume of 10x filtered yeast nitrogen base.

3.5.3.4   Yeast Display Titrations and Competition Binding

For cell surface titrations, EBY100 harboring the RBD display plasmid was grown in 1 ml M19D (5 g/l casamino acids, 40 g/l dextrose, 80 mM MES free acid, 50 mM citric acid, 50 mM phosphoric acid, 6.7 g/l yeast nitrogen base, adjusted to pH7 with 9M NaOH, 1M KOH) overnight at 30°C. Expression was induced by resuspending the M19D culture to OD_{600}=1 in M19G (5 g/l casamino acids, 40 g/l galactose, 80 mM MES free acid, 50 mM citric acid, 50 mM phosphoric acid, 6.7 g/l yeast nitrogen base, adjusted to pH7 with 9M NaOH, 1M KOH) and growing 22 h at 22°C with shaking at 300 rpm. For CR3022 IgG, yeast surface display titrations were performed as described by Chao et al. ([71]) with an incubation time of 4h at room temperature and using secondary labels anti-c-myc-FITC (Miltenyi Biotec) and Goat anti-Human IgG Fc PE conjugate (Invitrogen Catalog # 12-4998-82). Titrations were performed in biological replicates and technical triplicates (n = 6). The levels of display and binding were assessed by fluorescence measurements for FITC and SAPE using the Sony SH800 cell sorter equipped with a 70μm sorting chip and 488 nm laser.

To test the individual antibody panel binding to S RBD, EBY100 harboring the RBD display plasmid was grown from -80°C cell stocks in 1 ml SDCAA for 4h at 30°C. Expression was induced by resuspending the SDCAA culture to OD_{600}=1 in SGCAA and growing at 22h at 22°C with shaking at 300rpm. 1x10^5 yeast cells were labelled with 10 μg/ml antibody IgG for 30 min at room temperature in PBSF (PBS containing 1g/l BSA). The cells were centrifuged and washed with 200 μL PBSF. They were labeled with 0.6 μL FITC (Miltenyi Biotec), 0.25 μL Goat anti-Human IgG Fc PE (ThermoFisher Scientific) and 49.15 μL PBSF for 10min at 4°C. Cells were then centrifuged, washed with PBSF, and read on a flow cytometer to measure binding of the ACE2. Experiments were performed at least in biological duplicate.

Competitive binding assays on the yeast surface were performed between a free antibody and biotinylated ACE2. S. cerevisiae EBY100 harboring the RBD display plasmid was grown from -80°C cell stocks in 1 ml SDCAA for 4h at 30°C. Expression was induced by resuspending the SDCAA culture to $OD_{600}=1$ in SGCAA and growing at 22h at 22°C with shaking at 300rpm. $1x10^5$ yeast cells were labelled with 10 μg/ml antibody IgG for 30 min at room temperature in PBSF (PBS containing 1g/l BSA). The same cells were labelled with 30nM chemically biotinylated hACE2, in the same tube without washing, for 30min at room temperature in PBSF. The cells were centrifuged and washed with 200 μL PBSF. They were labeled with 0.6 μL FITC (Miltenyi Biotec), 0.25 μL SAPE (Invitrogen) and 49.15 μL PBSF for 10min at 4°C. Cells were then centrifuged, washed with PBSF. The pellet was resuspended in 100 μL and read on a flow cytometer to measure binding of the hACE2.

3.5.3.5   Yeast Display Screening of S and S RBD libraries

For full-length S ectodomain screening, pUC19-S-ecto-Nterm and pJS698 were independently linearized via digest with restriction enzymes at 37°C for 1 hour, and gel extracted based on size using Monarch DNA Gel Extraction Kit. The linearized regions were co-transformed in a molar ratio of 3:1 insert to vector into chemically competent EBY100 following published protocols ([90]). EBY100 cells were recovered in nuclease free water for 5 minutes and then plated on two different yeast media agar plates: SDCAA and M37D. Cells were incubated at 30°C for 3 days. After initial growth, colonies from each plate were selected and grown up at 30°C and 250 rpm overnight in the respective dextrose media: SDCAA, M37D. Cells were then induced in respective galactose media at an $OD_{6}00=1$ at three different temperatures, 18°C, 22°C, and 30°C for 20 hours.

Induced EBY100 cells were washed with PBSF (8 g/L NaCl, 0.2 g/L KCl, 1.44g/L Na_2HPO_4, 0.24g/L KH_2PO_4, and 1g/L bovine serum albumin, pH to 7.4 and filter sterilized) and resuspended in PBSF at an $OD_{6}00=10$. The cells were then incubated with either 500nM of the biotinylated

ACE2-Fc or 500nM of the biotinylated CR3022 for 1 hour at room temperature. The cells were then washed with PBSF and labeled with anti-c-myc fluorescein isothiocyanate (FITC) (Miltenyi Biotec) and streptavidin–R-phycoerythrin (SAPE) (Invitrogen) and incubated on ice for 10 minutes.

The Spike mutagenic library was labeled with CR3022 and, separately, ACE2-Fc under the optimal conditions were screened. Approximately 108 yeast cells were sorted using fluorescence activated cell sorting (FACS), and the top 1% of cells by fluorescence were collected. The two resulting sorted libraries were expanded and sorted in a second round, again screening 108 cells and collecting the top 1% by fluorescence intensity. The selected populations were amplified and purified based on tile, deep sequenced, and count data compared with a reference population.

For the escape mutant screening of the S RBD, $3x10^7$ induced EBY100 yeast cells displaying S RBD were labelled with 10 μg/ml antibody IgG for 30 min at room temperature with mixing by pipetting every 10 min in PBSF (PBS containing 1g/l BSA). The same cells were labelled with 75nM chemically biotinylated ACE2, in the same tube, for 30min at room temperature in PBSF with mixing by pipetting every 10 min. The cells were centrifuged and washed with 1mL PBSF. Cells were then labeled with 1.2 μL FITC, 0.5 μL SAPE and 98.3 μL PBSF for 10min at 4°C. Cells were centrifuged, washed with 1mL PBSF, resuspended to 1 mL PBSF and sorted using FACS. Multiple gates were used for sorting as shown in **Figure C.7**, including an FSC/SSC$^+$ gate for isolation of yeast cells, FSC-H/FSC-A gate to discriminate single cells, a FSC-A/FITC$^+$ gate selects the cells displaying the RBD on their surface and from this last gate, the top 2% by a PE$^+$/FITC$^+$ is collected. At least $2.0x10^5$ cells were collected and were recovered in SDCAA with 50 μg/mL Kanamycin and 1x PenStrep for 30h. For the biological replicates (**Figure C.14**) the ACE2 concentration was 30nM but all other conditions were identical.

3.5.3.6   Deep Sequencing Preparation

Libraries were prepared for deep sequencing following the "Method B" protocol from Kowalsky et al ([61]) exactly as described for the spike ectodomain libraries and with a few changes for the RBD libraries. A Monarch PCR & DNA Cleanup kit was used. PCR of extracted and cleaned-up yeast plasmid DNA was performed using 2xQ5 HotStart Master Mix (NEB) and the following protocol:

98°C	1min	
98°C	10s	
64°C	20s	25 cycles
72°C	30s (replicate 1) or 1min (replicate 2)	
72°C	2min	
4°C	Hold	

Primers used in library prep are given in Key resources table. Amplicons were fractionated by agarose gel electrophoresis and purified using a Monarch DNA Gel Extraction Kit (NEB). Samples were then further purified using Agencourt Ampure XP beads (Beckman Coulter), quantified using PicoGreen (ThermoFisher), pooled, and sequenced on an Illumina MiSeq using 2 x 250 bp paired-end reads at the BioFrontiers Sequencing Core (University of Colorado, Boulder, CO).

3.5.3.7   Molecular Dynamics Simulations

GROMACS 2018.3 ([95]) was employed for all molecular dynamics (MD) simulations along with the TIP3P ([96]) water model and Amber99SB-ILDN ([97]) force field to model the receptor binding domain (RBD) of the spike (S) protein of SARS-CoV-2 and neutralizing antibodies CC12.1 and CC12.3. Simulations were initiated from crystal structures of the RBD in complex with CC12.1

(PDB code 6XC2 ([37]) and CC12.3 (PDB code 6XC4 ([37]). All systems containing a positive charge were neutralized by the addition of Cl- ions, also modeled with the Amber99SB-ILDN force field. Each simulation consisted of approximately 192,000 atoms.

A steepest descent energy minimization of the initial coordinates for each system was carried out for 5,000 steps. NVT equilibration simulations were then performed for 0.5 ns at 310 K with the Bussi-Donadio-Parrinello ([98]) thermostat. Subsequent NPT equilibration simulations were performed for 1 ns at 310 K and 1.0 bar, using the same thermostat and Berendsen ([99]) barostat. The time constant for coupling in both the NVT and NPT simulations was 0.1 ps. Production simulations in the NPT ensemble were then carried out at 310 K and 1.0 bar with the Bussi-Donadio-Parrinello thermostat and Parrinello-Rahman ([100]) barostat. Long-range electrostatic interactions were calculated using particle mesh Ewald summations and a cutoff of 1.0 nm, and Lennard Jones interactions were calculated over 1.0 nm and shifted beyond this distance. Neighbor lists were updated every 10 steps with a cutoff of 1.0 nm. Bonds between hydrogen and heavy atoms were constrained with the LINCS ([101]) algorithm. Furthermore, periodic boundary conditions were used in all simulations in all directions. Production simulations were carried out for 100 ns, leading to a total of 0.8 microseconds of simulation time across the eight simulations.

3.5.3.8   Pseudo Neutralization Assays

SARS-CoV-2 pseudovirus neutralization assays were performed as previously described ([4]). Briefly, pseudovirus was generated by cotransfecting MLV-gag/pol and MLV-CMV-Luciferase plasmids with truncated wildtype SARS-CoV-2 or mutant SARS-CoV-2 plasmid respectively onto HEK293T cells. After 48h or 72h of transfection, supernatants containing pseudovirus were collected and frozen at -80°C. Neutralization assay was performed as follows. First, monoclonal antibodies were serially diluted into half-area 96-well plates (Corning, 3688) and incubated with pseudovirus at 37°C for 1 h. Next, HeLa-hACE2 cells were transferred in the 96-well plates at

10,000 cells/well. After 48h of incubation, supernatants were removed, cells were lysed with 1x luciferase lysis buffer (25 mM Gly-Gly pH 7.8, 15 mM $MgSO_4$, 4 mM EGTA, 1% Triton X-100). Finally, Bright-Glo (Promega, PR-E2620) was added onto 96-well plates according to manufacturer's instructions. Neutralization IC50s were calculated using "One-Site LogIC50" regression in GraphPad Prism 8.0. Pseudovirus mutant constructs were generated by amplifying two overlapped fragments of SARS-CoV-2 mutant sequences with Q5 enzyme (NEB, M0492) following manufacturer's instructions. Two fragments were then joint into one fragment by bridge PCR, and gibson cloned into digested pcDNA3.3 backbone.

3.5.4   Quantification and statistical analysis

3.5.4.1   Dissociation constants

The dissociation constants on Figure 2D represent the mean of the replicates (values show as open circles). There are two replicates for each single point mutant and 4 for the wilt type.

3.5.4.2   Deep Sequencing Analysis

All deep sequencing data analysis was performed by scripts written in Python, available at GitHub (`https://github.com/WhiteheadGroup/SpikeRBDStabilization.git`).

Because all sequenced samples were PCR amplicons of known length, paired-end reads were merged by aligning at the known overlap. Mismatches in overlapping regions were resolved by selecting the base pair with the higher quality score and assigning it a quality score given by the absolute difference of the quality scores at the mismatch. Paired reads with more than 10 mismatches in the overlapping region and merged reads containing any quality score less than 10 were discarded. The total number of retained reads in each sample was recorded as n_i, the number of reads in sample i.

Each read was compared to the wild-type sequence to identify all mutations. Counts for

synonymous single mutations were combined to give k_{ij}, the number of reads in sample i encoding the single amino acid mutation j. Reads including multiple mutations or mutations not encoded in the library oligos were not analyzed further. The frequency of single mutant j in sample i was calculated as $f_{ij} = k_{ij}/n_i$.

Each experiment consisted of two samples: a reference sample r and a selected sample s. For each experiment, the risk ratio of variant j was calculated as $\rho_j = f_{sj}/f_{rj}$ i.e. the ratio of the variant's frequency in the selected population to its frequency in the reference population. Enrichment ratios were calculated as the binary logarithm of the risk ratio: $ER_j = \log_2 \rho_j$. Variants with five or fewer counts in the reference population were not analyzed further. Variants with at least five counts in the reference population but no counts in the selected population were given a pseudocount of one.

3.5.4.3   Determining hits from yeast display screens

For each escape mutant screen, we collected the top 2% (PE channel) of the population of FITC$^+$ (RBD displaying) cells. This population was not labeled with biotinylated ACE2 and so serves as a null experiment where the observed enrichment ratios are due to other sources of variance and not to differential nAb binding. We fit the distribution of enrichment ratios for each of these control samples using kernel density estimation (KDE) (SciPy's $scipy.stats.gaussian_kde$ with default parameters) ([102]). We then treated this distribution estimate as an empirical null hypothesis. Under this null hypothesis, we expect $N(1 - F(ER_t))$ false positives, where N is the number of variants tested, F is the cumulative distribution function (CDF) of the control ER KDE, and ER_t is a threshold. Therefore, for a target false discovery rate (FDR), we chose $ER_t = F^{-1}(1 - FDR/N)$, where F^{-1} is the inverse CDF of the KDE. In data from samples labeled with nAbs, we then tested the hypothesis that each observed ER was greater than the associated ER_t using an one-sided exact Poisson rate ratio test ($statsmodels.stats.rates.test_poisson_2indep$ from

the Python library statsmodels) ([103]). For these tests, the null ratio was 2^{ER_t}. The counts were given by the number of reads for the variant in the selected and reference populations, respectively, and the exposures were given by the total number of reads in the reference and selected populations, respectively. For this analysis, we identified hits for replicate 1 (tiles 1 & 2 for nAbs CC6.29, CC12.1, and CC12.3) using a target FDR of 1 and a Poisson rate ratio test significance level of 0.01. For replicate 2 (tile 2 for nAbs CC6.31, CC12.13) escape mutant hits were identified using a target FDR of 1.

For the full-length S ectodomain screen, our null experiment was the collected reference populations without selections for each of the ACE2-Fc and CR3022 experiments. These reference populations were passaged, sorted, and amplified identically to the sorted libraries except that no screen was employed. We fit the distribution of enrichment ratios for these control samples using a logistic CDF (custom MATLAB script), and the empirical FDR was calculated exactly as above.

3.6   Acknowledgements

3.7   Author Contributions

Designed experiments: IMFU, PJS, MBK, ERR, KGS, JJ, TAW; Performed experiments: IMFU, PJS, MBK, SB, LP, FZ, ERR, ACL; Performed simulations: ERR, KGS; Developed algorithms and software: PJS, CMH, TAW; Wrote paper: IMFU, PJS, MBK, KGS, TAW.

Chapter 4

Using deep mutational scanning to inform the engineering SARS-CoV-2 Broadly neutralizing antibodies

This chapter has been adapted from Zhao et al. 2022 iScience[5]

The rapid spread of SARS-CoV-2 variants poses a constant threat of escape from monoclonal antibody and vaccine countermeasures. Mutations in the ACE2 receptor binding site on the surface S protein have been shown to disrupt antibody binding and prevent viral neutralization. Here, we used a directed evolution-based approach to engineer three neutralizing antibodies for enhanced binding to S protein. The engineered antibodies showed increased in vitro functional activity in terms of neutralization potency and/or breadth of neutralization against viral variants. Deep mutational scanning revealed that higher binding affinity reduces the total number of viral escape mutations. These data suggest that monoclonal antibodies for antiviral indications would benefit from affinity maturation to reduce viral escape pathways.

4.1 Introduction

Since 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has had devastating consequences for global health and economies. Vaccines and therapeutics were quickly developed to combat the disease, with several highly effective vaccines being developed that elicit immune responses against the SARS-CoV-2 spike (S) trimer [104]. Neutralizing antibodies (nAbs) have been identified that target several distinct epitopes on the S trimer, but the majority of nAbs target the receptor binding domain (RBD) [50, 105]. While vaccines are crucial for controlling COVID-19, recombinant nAbs can be used to supplement vaccination efforts. They offer potential benefits for populations that respond poorly to vaccines, such as immunocompromised individuals, as a post-exposure preventive measure and as a therapeutic option to prevent hospitalization [106, 107].

One of the unique challenges in using a neutralizing monoclonal antibody (mAb) for antiviral indications is addressing existing viral diversity and the high mutational propensity in viruses that can give rise to resistant viral variants. Since the discovery of SARS-CoV-2 in 2019, thousands of viral variants containing synonymous and non-synonymous mutations have been documented (`https://nextstrain.org/ncov/gisaid/global`). A growing number of these new variants (termed 'Variants of Interest' or VOIs, and 'Variants of Concern' or VOCs) contain mutations that increase infectivity and/or allow viral escape from monoclonal nAbs elicited against the original SARS-CoV-2 [78, 108, 109, 110].

To reduce the likelihood of viral escape from nAbs, several strategies are in use. The most direct mitigation of escape is to develop antibodies that target a functionally important and conserved region, hence reducing the number of mutations that can allow viral escape without incurring a fitness cost [111, 112, 113]. Another approach is to use use cocktails of at least two nAbs targeting different epitopes, so multiple mutations are necessary for viral escape [114, 66, 115, 116]. A third approach that is less well explored is to in vitro affinity mature the nAb against the target antigen

to increase the binding affinity, helping to mitigate the impact of the viral mutations [47]. Here, we explore how increased binding affinity impacts the in vitro neutralization breadth and potency of three COVID-19 nAbs, CC12.1, CC6.30 and CC6.33 [4], and how these affinity improvements impact the in vivo protective capability of these nAbs.

4.2   Results

4.2.1   Structural analysis of nAbs CC6.30 and CC6.33

To better understand the molecular contacts of antibodies, we used cryo-electron microscopy (cryoEM) to solve the structures of CC6.30 and CC6.33 of the antibodies in complex with stabilized SARS-CoV-2 S trimers. The structure of CC12.1 was previously reported to bind Class 1 epitope and directly compete with ACE2 [50, 117, 110]. nAb CC6.33 targets the Class 2 epitope site with contributions from all heavy chain (HC) and light chain (LC) complementarity determining regions (CDRs) except for LCDR2 and it only binds the RBD-up conformation. CC6.33 binds a non-overlapping epitope to that of CC6.30, with the HC and LC interface centered on the N343 glycan. This nAb binds the RBD-down conformation and the most stable reconstruction has 2 down RBDs bound by the antibody.

4.2.2   Engineering higher affinity SARS-CoV-2 antibodies

Next, using the SAMPLER strategy developed by [118], the three antibodies were affinity matured by generating libraires that contained one mutation per CDR loop from the starting sequence. The libraries were displayed on the surface of yeast and iterative rounds of selections were used to enrich for clones with higher affinity for SARS-CoV-2 RBD (for CC12.1 and CC6.30 libraries) or S protein (for the CC6.33 library). The enriched clones were then combined into a heavy/light combinatorial library and screened again with the same four-round selection strategy to identify the optimal heavy/light pairs [118]. 12 improved variants of each antibody were selected to be reformatted and expressed as IgG for characterization. All enhanced (e) eCC12.1, eCC6.30,

and eCC6.33 variants bound SARS-CoV-2 RBD with a higher affinity than the parental antibodies.

4.2.3   Neutralizing activity of engineered antibodies

We next sought to investigate how the evolving viral diversity of SARS-CoV-2 variants impacts the binding and the neutralization function of our parental and select engineered nAbs. CC12.1 neutralized Alpha, Delta, and Kappa variants with an IC_{50} comparable to Wuhan-1 virus, however, Beta and Gamma VOIs completely escaped from this nAb. Analysis of the individual variants found that K417N (from Beta VOI) and K417T (from Gamma VOI) facilitated this escape, consistent with the previous observation that most VH3-53-class nAbs are sensitive to these mutations [3, 110, 119]. 11 out of 12 mAbs neutralized the Gamma lineage and 9 out of 12 neutralized the Beta lineage. Importantly, eCC12.1.4 and eCC12.1.7 neutralized all VOCs, including Beta and Delta, containing the K417N/T mutations.

CC6.30 and eCC6.30.2 were effective against the Alpha variant, showed significantly reduced function against the Delta variant, and were completely unable to neutralize the Beta, Gamma or Kappa variants. Finally, the parental CC6.33 and eCC6.33.8 were effective at neutralizing all variants tested with similar IC_{50}s to the original Wuhan-1 SARS-CoV-2, consistent with the observation that CC6.33 recognizes the conserved class 3 epitope.

4.2.4   Mapping RBD escape mutations for CC12.1 and eCC12.1.4

We next asked whether engineering high affinity nAbs reduces the pathways for viral escape compared with the parental nAbs or if they simply shifted the escape mutations to other positions. Deep mutational scanning libraries of RBD were generated and used to determine the mutations on RBD that prevented CC12.1 and eCC12.1.4 from blocking ACE2 binding in vitro (**Figure C.18A**) (Francino-Urdaniz et al., 2021). 94.5% (2250/2380) of all possible single mutations were scanned (**Table D.6**). Consistent with previous reports [3] and our neutralization screening, CC12.1 is vulnerable to multiple mutations at K417 with a false discovery rate (FDR) below 0.1 for K417N/T
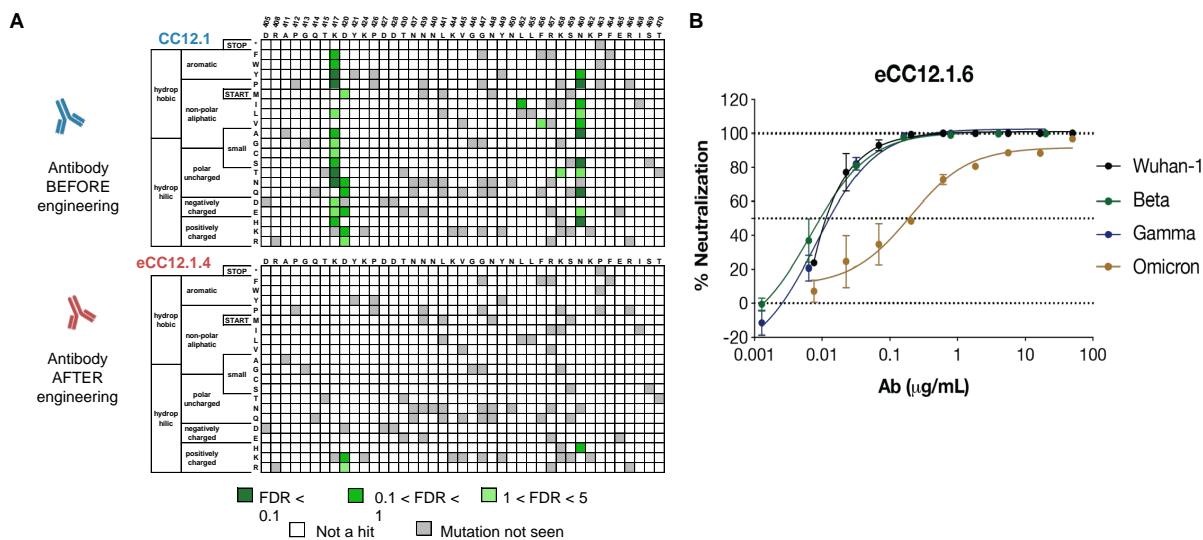
Figure 4.1: **Identification of escape mutants and validation of broadly neutralizing antibodies. A.** Heatmaps of the antibody CC12.1 before (top) and after (bottom) engineering. **B.** Pseudo-neutralization assay of the engineered antibody eCC12.1.6 against different variants of SARS-CoV-2. x-axis represents the antibody concentration in μg/ml and y-axis represents the % neutralization of the antibody.

but eCC12.1.4 is able to accommodate all mutations at K417 (**Figures 4.1A, and C.18B**). We also detected multiple mutations at position D420 and N460 that confer escape from the parental CC12.1 (**Figures 4.1A**). Alanine scanning had identified these D420 and N460 residues as important for public VH3-53 SARS-CoV-2 antibodies [120, 119], but structural analysis shows these two positions on the periphery of the CC12.1 epitope and making relatively insignificant contacts to the antibody.

Pseudoviruses containing the individual D420K, N460H, N460P, and N460A mutations, identified as potential escape mutations in the deep mutational scanning, were produced and evaluated to determine if parental CC12.1 and several eCC12.1 variants were sensitive to these mutations in a neutralization assay. In agreement with RBD library screening, a D420K substitution completely disrupted CC12.1 neutralization, whereas substitutions at the N460 residue significantly decreased its neutralization potency by 12- to 246-fold . By contrast, although D420K and N460H were identified as potential escape mutations against eCC12.1.4, neutralization potency was reduced by a

more modest 8-fold against D420K and remained insensitive to a N460H substitution . These data suggest that increasing the affinity of SARS-CoV-2 nAbs restricts the potential escape mutations that can arise in RBD, rather than just altering the critical nAb contacts and shifting the escape mutations to a comparable number of different positions and/or mutations. This is particularly important in the context of developing antiviral antibodies where viral escape is a serious and constant threat.

4.2.5   Omicron neutralization

During the preparation of this manuscript the Omicron VOC was reported. Omicron completely escaped from the parental CC6.30 and all eCC6.30 variants as well as the parental CC6.33 and all eCC6.33 variants. However, we observed that Omicron was resistant to parental CC12.1 and 11 of 12 eCC12.1 variants, however, eCC12.1.6 retained neutralizing activity against Omicron (IC$_{50}$ of 0.20 μg/mL), albeit with approximately 25-fold reduced potency (**Figure 4.1B**).

4.3   Discussion

The emerging SARS-CoV-2 variants have challenged vaccine-induced immunity and therapeutic mAbs, and functional nAbs with prophylactic and therapeutic efficacy against VOCs are desired. To this end, investigators have searched for nAbs that are largely resistant to viral antigenic variability by screening for nAbs that broadly neutralize SARS-CoV-2 variants and/or other sarbecoviruses. Here, we explored an alternative and complementary strategy of engineering nAbs to have higher affinity for their target antigen through directed evolution. Affinity improvements do help to expand the breadth of antibody reactivity, allowing them to better neutralize variants that contain mutations in and around the antibody epitope. This was particularly evident with the eCC12.1 nAbs that are part of the shared VH3-53 lineage and are broadly susceptible to the K417T/N mutations found in Beta and Gamma VOCs [3, 110].

Our saturated mutagenesis screening showed that the affinity maturation restricted the num-

ber of potential escape mutations rather than just altering them to different positions. Although increasing the affinity can restrict escape mutations, it does not abrogate them entirely, as eCC12.1 variants showed modest sensitivity to the D420K mutation, and all eCC6.30 variants were still unable to bind or neutralize the E484K mutation found in Beta and Gamma VOIs. In the extreme case of the Omicron variant that contains so many mutations on the RBD, especially within the footprint of class 1 antibodies, there was still one eCC12.1 variant that had significant neutralizing activity. This finding illustrates the value of affinity maturation in the context of natural infection where the generation of a diverse set of related antibodies, as was done here in vitro, will likely generate some antibodies able to bind to and act against many different viral variants, including those with multiple mutations as for Omicron. It is possible that part of the reason so many clinical antibody candidates failed against Omicron is that most were selected shortly after COVID infection before much affinity maturation had occurred [113].

4.4 Methods

Full methods is available online.

4.4.1 RBD library generation and identification of escape mutants

Yeast display plasmids pJS697 and pJS699 used for surface display of Wuhan-Hu-1 S RBD N343Q were previously described [2]. Using these plasmids, 119 surface exposed positions on the original Wuhan-Hu-1 S RBD N343Q (positions 333–537) were mutated to every other amino acid plus stop codon using degenerate NNK primers using comprehensive nicking mutagenesis [82] exactly as previously described [3, 121]. Two tiles were constructed for compatibility with 250bp paired end Illumina sequencing (tile 1: positions 333–436; tile 2: positions 437–527). Libraries were transformed into *S. cerevisiae* EBY100 and stocks of $1e^8$ viable yeast cells in 1mL were stored in yeast storage buffer (20w/v % glycerol, 200 mM NaCl, 20 mM HEPES pH 7.5) at -80°C. Library coverage was confirmed by 250 bp paired end Illumina deep sequencing, with statistics reported in **Table D.6**.

S RBD escape mutants are identified by a competitive assay between a nAb and soluble ACE2 as fully described in Francino-Urdaniz et al. (2021)[3]. Briefly, yeast cells are grown in SDCAA for 4h at 30°C, pelleted, and then induced in SGCAA for 22h at 22°C. Cells are washed thoroughly in PBSA (PBS containing 1g/L BSA) and then 3×10^7 induced EBY100 yeast cells displaying S RBD were labeled with 10 μg/mL nAb IgG for 30 min at room temperature with mixing by pipetting every 10 min in PBSA. The same cells were labeled with 75 nM chemically biotinylated ACE2, in the same tube, for 30 min at room temperature in PBSA with mixing by pipetting every 10 min. The cells were centrifuged and washed with 1mL PBSA. Cells were then labeled with 1.2 μL FITC,0.5 μL SAPE and 98.3 μL PBSA for 10 min at 4°C. Cells were centrifuged, washed with 1mL PBSF, resuspended to 1 mL PBSA and sorted using FACS. Two additional populations were sorted: a reference population containing only an FSC/SSC gate for isolation of yeast cells (see below) and a control population of library not competitively labeled.

To discriminate cell populations FACS gates are used as shown in **Figure C.18**: an FSC/SSC gate for isolation of yeast cells, FSC-H/FSC-A gate to discriminate single cells, a FSC-A/FITC$^+$ gate selects the cells displaying the RBD on their surface and the top 2% of cells by a PE$^+$/FITC$^+$ gate is collected. At least 2.0×10^5 cells are collected and recovered in SDCAA with 50 μg/mL Kanamycin and 1$\times$ Pen/Strep for 30 h at 30°C. The collected DNA is sequenced using 250 bp PE on an Illumina MiSeq and analyzed with the code developed by Francino-Urdaniz et al. (2021) [3]. Outputs from the code are a per-mutation enrichment ratio defined as the log2-transform of the change in frequency of the selected population relative to the reference population and a false discovery rate (FDR) as previously described [3].

4.5   Acknowledgements

4.6   Author contributions

F.Z. and J.G.J. conceived and designed the study. F.Z. and S.B. performed yeast library preparation, yeast cell staining, sorting, sequencing, and cloning experiments. F.Z., S.B., A.B., and O.L. expressed and purified the monoclonal antibodies. G.O. and H.T. performed Cryo-EM sample preparation, imaging, data processing and model building. F.Z., S.B., and A.B. character-ized monoclonal antibodies in functional assays and biophysical analysis. A.B. and O.L. generated mutant RBD protein plasmids and expressed re- combinant RBD and S proteins. J.W. performed SPR experiment and analysis. A.B., S.B., and L.P. generated pseudovirus mutant constructs, pro-duced pseudovirus and performed neutralization assay. I.F.U and T.A.W. generated RBD mutant library, analyzed and identified escaped mutants. C.K. and N.S. performed hamster protection experiments including antibody infusion, virus challenge, weight analysis, viral load measurement. F.Z., C.K., G.O., D.R.B., and J.G.J. wrote the manuscript and all authors reviewed and edited the manuscript.

Chapter 5

Benchmarking cassette-based deep mutagenesis by Golden Gate assembly

Protocols for the construction of large, deeply mutagenized protein encoding libraries via Golden Gate assembly of synthetic DNA cassettes employ disparate, system specific methodology. Here we benchmark a broadly applicable Golden Gate method for building user-defined libraries. We demonstrate that a 25 μl reaction, using 40 fmol of input DNA, can generate a library on the order of 1×10^6 members and that reaction volume or input DNA concentration can be scaled up with no losses in transformation efficiency. Such libraries can be constructed from dsDNA cassettes generated either by degenerate oligonucleotides or oligo pools. We demonstrate its real-world effectiveness by building custom, user-defined libraries on the order of 10^4 to 10^7 unique protein encoding variants for two orthogonal protein engineering systems. We include a detailed protocol and provide several general-use destination vectors.

5.1   Introduction

Cassette assembly has become a powerful way to create protein libraries thanks to modern synthesis technologies that can quickly and affordably produce custom DNA fragments [122, 123, 92, 124]. Some of the most useful tools for assembling such fragments are Type IIs enzymes, which cut outside of their DNA recognition site and leave a user-defined four base pair overhang. While labs have employed these enzymes to manipulate DNA for the last several decades [125, 126] a major development in their use came in 2008 when Engler et al. established a cloning method using the Type IIs enzyme BsaI, which they called Golden Gate assembly [127]. Since then, the method's popularity has grown [128, 129, 130], due in part to its ability to connect fragments in a specific order and without a restriction scar[131].

Although the assembly of synthetic DNA by Golden Gate has clear utility in building large protein libraries, a generalized procedure has yet to be established. While several labs have used Golden Gate to build libraries from cDNA and synthetic DNA fragments, their assembly protocols and resulting efficiencies varied widely and often contained time-consuming, system-specific steps [132, 133, 134, 135, 136]. In 2019 Püllmann et al. developed a more general Golden Gate protocol for creating site-saturation libraries from oligonucleotides, which was well characterized and included a useful script for primer design [137]. However, the base version of this protocol was only shown to generate a single point mutation and a more complicated protocol in which subcloning was needed to make a library of 60 variants.

Here we provide a simple, broadly applicable Golden Gate procedure that can be used to build large ($>10^7$), site-specific, deeply mutagenized libraries. We present data benchmarking the procedure's efficiency under different use conditions and demonstrate its effectiveness in constructing libraries from mixed base-containing oligonucleotides and custom synthesized oligo pools. We also provide a detailed protocol (**Appendix B** with discussion of important design considerations and three general-use destination vectors deposited on Addgene.

5.2 Results/Discussion

5.2.1 A standardized protocol for library generation

The general workflow for library generation using our protocol involves design and creation of a destination vector and mutagenic cassettes, assembly via a Golden Gate reaction, and transformation into bacteria. The destination vector must be designed and created with a selection marker, BsaI sites, and specific overhangs such that cassette(s) introduction results in reconstitution of a full-length gene encoding sequence, and replacement of the marker which allows rapid assessment of incorporation efficiencies (**Figure 5.1A**). In parallel, cassettes are designed with BsaI sites and overhangs arranged for sequential insertion into the destination vector (**Figure 5.1A**) using subsets of the high-fidelity overhangs outlined by Potapov et al. to reduce inefficiencies from imperfect ligation[87]. Subsequently, the destination vector and cassettes are mixed with BsaI-HFv2 and T4 ligase and PCR cycled between 37°C and 16°C, during which time the selection marker and cassette ends are removed, the cassettes anneal with the vector, and the annealed DNA is ligated.

We developed a modified Golden Gate protocol that allowed us to rapidly assemble libraries while maintaining high numbers of transformants. For the base version of this protocol, we used a 25 μ l reaction containing 40 fmol of the destination vector and each cassette. Additionally, our reactions were PCR cycled for 60 cycles with each step only lasting one minute, as described by Strawn et al. [138], allowing the Golden Gate reaction, and bacterial transformation to be performed in under 8 hours (**Figure 5.1A**). We have included a protocol describing the design and creation of libraries using this technique in more detail (**Appendix B**). Using this protocol, with properly designed cassettes and a destination vector in hand, a new library for any protein can be generated in a single day. To facilitate this, we built general use destination vectors for yeast surface display (pND003), MBP-tagged protein expression in E. coli (pND004), and yeast two-hybrid assays (pND005), each of which contain a GFP marker and BsaI sites with high fidelity overhangs. To confirm their functionality, a Golden Gate reaction was performed with pND003 and cassettes

that encode for a monomeric variant of the plant abscisic acid receptor PYR1 (H60P, N90S) [139]. Comparison of plates from a transformation of pND003 alone and a transformation of the Golden Gate reaction confirmed cassette insertion and demonstrated that removal of the GFP marker allowed for assessment of incorporation efficiencies (**Figure 5.1B**). We used numbers of GFP negative colonies to calculate incorporation percentages for all Golden Gate reactions described in this paper with a median incorporation of 99.7% (range 81% to >99.9%, n=40) (**Figure 5.1C**; **Table D.4**).

To verify that our protocol would result in genes encoding full-length, functional protein, we assessed the ability of an SARS-CoV-2 Omicron chimeric S RBD to bind ACE2 and the neutralizing monoclonal antibody CC12.1 [4, 3]. For these experiments we created a destination vector with the coding sequence for the Omicron chimeric RBD (SARS-CoV-2 S RBD (333-541) Wuhan-1 with mutations S477N, E484A, Q498R, N501Y and Y505H) and the corresponding dsDNA coding cassettes, assembled them into the RBD destination vector using our Golden Gate protocol, transformed the resulting plasmids into yeast, and expressed the isogenic Golden Gate-derived RBDs (GG RBD) on the surface of yeast. As a positive control, we also expressed the previously described Wuhan-1 S RBD N343Q (Wuhan-1) [3] on the surface of yeast. We then compared the ability of the GG RBD and Wuhan-1 to bind Fc-ACE2 and CC12.1 (which contains an Fc) using flow cytometry. When labeled with an anti-Fc PE, we observed specific PE fluorescence resulting from binding to ACE2 or CC12.1 for both GG RBD and Wuhan-1 (**Figure 5.1D**). We then tested GG RBD functionality in a similar manner for 16 different yeast colonies from four separate Golden Gate reactions and saw consistent levels of CC12.1 binding for all 16 variants, highlighting the ability of this protocol to reproducibly assemble full-length sequences (**Figure 5.1E**).

5.2.2   Benchmarking a scalable single day Golden Gate library generation protocol

Next, we benchmarked the number of transformants that can be generated using this protocol as a function of cassette number, reaction size, and input DNA. For all these experiments we used
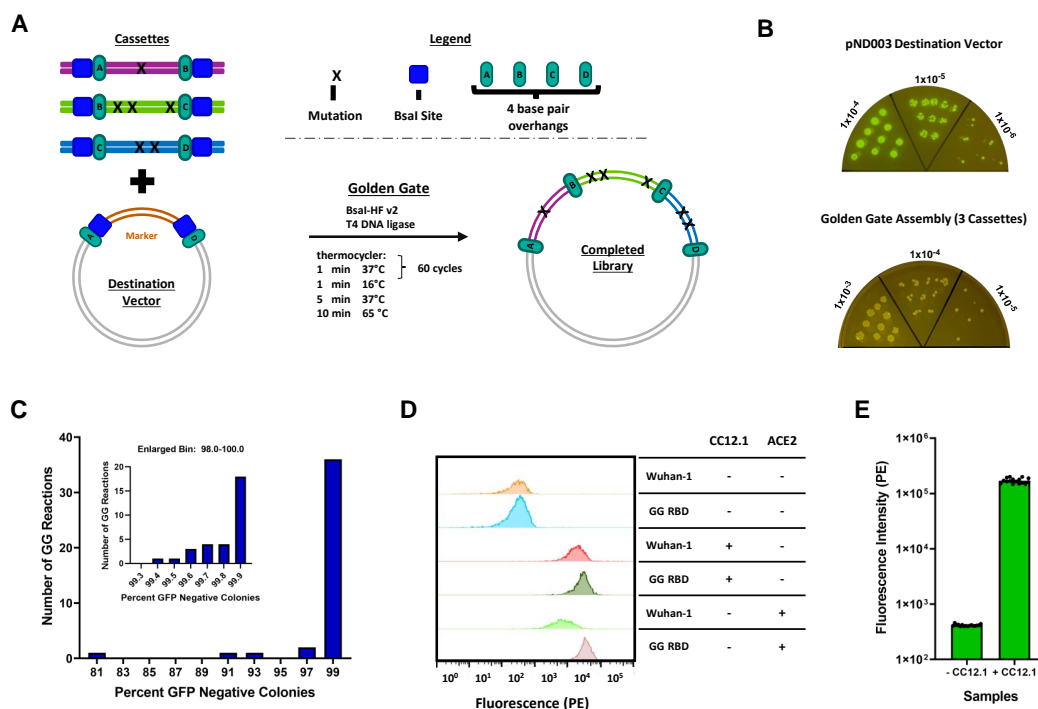
Figure 5.1: **Establishing the functionality of a Golden Gate protocol for library generation.** **A.** Schematic showing library generation by assembly of mutagenic cassettes into a destination vector via Golden Gate. **B.** Dilution plating of E.coli transformations with 40 fmol of our destination vector pND003 (left) and a Golden Gate reaction performed with 40 fmol of both pND003 and three wild type PYR1 cassettes (right). Listed numbers represent fold dilution. **C.** Histogram showing the percent cassette incorporation for all performed Golden Gate reactions. Inset shows all reactions from the 98-100% bin redistributed in bins with 0.1% intervals. Incorporation percentages were calculated by comparing the number of green (GFP) and white (non-GFP) colonies on dilution plates after transformation. **D.** Functional comparison of GG RBD and Wuhan-1 RBD using yeast surface display. RBD displaying yeast were incubated in the absence (-) or presence (+) of saturating concentrations of PE-labeled CC12.1 (antibody) or ACE2, followed by assessment of binding by flow cytometry. **E.** Assessment of CC12.1 binding for 16 colonies displaying RBD from 4 different Golden Gate reactions. Individual values represent mean PE fluorescence intensity of RBD-displaying populations.

our pND003 vector and PYR1 cassettes. Our baseline for all three experiments was a 25 μl reaction containing 40 fmol of the destination vector and a single PYR1 cassette, which generally results in $7 - 8 \times 10^5$ transformants and >96% cassette incorporation (**Figure 5.2**, **Table D.4**).

Since previous studies have noted decreasing Golden Gate efficiencies when increasing the number of 'parts' (number of cassettes plus destination vector) [133, 140], we sought to quantitate

the number of transformants as a function of the number of input cassettes. For this experiment, we performed four reactions with one to four PYR1 cassettes under conditions that were otherwise identical to the baseline reaction. We found that transformation efficiencies were modestly affected by increasing cassette number, with a four-cassette (five-part) assembly showing a minor reduction in transformants (3.7×10^5, n = 2) over a one-cassette (two-part) assembly (8.4×10^5, n = 2) (**Figure 5.2A**, **Table D.4**). Thus, increasing cassette numbers should not hinder most library designs.

We then assessed the ability of our protocol to be scaled by total reaction size by comparing our baseline 25 μl reaction with 100 μl and 200 μl reactions. We found that the number of transformants scaled at least linearly, with the 200 μl reaction (8x volume) generating an approx. 24-fold increase in transformants with no loss in incorporation efficiency (p-value 0.03; **Figure 5.2B**, **Table D.4**). We speculate that this trend results from decreased relative DNA loss while working with small constant volumes during the PCR cleanup and transformation steps. Since a 100 μl reaction can be performed in a single PCR tube, several orders of magnitude higher numbers of transformants should easily be achieved by pooling multiple Golden Gate reactions in a single PCR cleanup column.

Subsequently we tested scaling of DNA concentration by comparing our baseline reaction with reactions in which the amount of destination vector and cassette DNA was increased to 80 and 160 fmol, without increasing reaction size. We again observed a linear increase in transformants, with 160 fmol (4x increase) of input DNA resulting in a 19-fold increase in transformants (p value 0.13 for whether the 160 fmol reaction gives more than a 4x increase in number of transformants) and no loss in incorporation efficiency (**Figure 5.2C**, **Table D.4**). Thus, reactions can be scaled by both volume and DNA concentration.
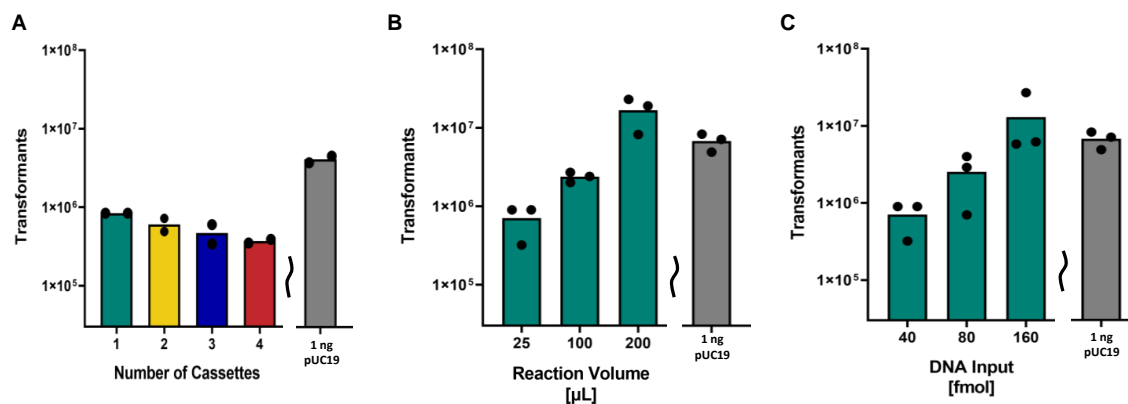
Figure 5.2: **Benchmarking a Golden Gate protocol for library generation A.** Transformation efficiencies for Golden Gate reactions performed with differing numbers of cassettes. 25 μl reactions were performed in duplicate with 40 fmol of both pND003 and one, two, three, or four wild type PYR1 cassettes. Numbers of transformants were calculated by plating serial dilutions of the recovered cells. pUC19 is shown as a control reaction for assessing efficiency of electrocompetent cells. **B.,C.:** Transformation efficiencies as a function of Golden Gate reaction size (B) or input DNA (C). A baseline 25 μl reaction using 40 fmol of both pND003 and one wild type PYR1 cassette was compared with reactions having increased size or input DNA. Reactions were performed in triplicate. Numbers of transformants were calculated by plating serial dilutions of the recovered cells. (**B.**) All reaction components were scaled 1:1, including DNA concentration, with increasing volume. (**C.**) The total amount of input DNA for the destination vector and cassette were increased while all other reaction conditions were held constant.

5.2.3 Creation of a site-specific protein libraries using both DNA Ultramers and oligo pools

We next assessed the ability of our protocol to produce complete libraries from different types of synthetic DNA by building libraries for the Omicron chimeric RBD. For this we designed three site-specific combinatorial libraries, all covering the same 110 contiguous positions, assembled using a combination of three mutant cassettes (**Figure 5.3A**).

We generated cassettes from synthetic dsDNA (eBlocks), mixed-base degenerate long oligonucleotides (Ultramers), and ssDNA sourced from custom oligo pools (**Figure 5.3B**). Generally, we used eBlocks to encode unmutated regions of protein, while Ultramers and oligo pools were used as mutagenic cassettes. dsDNA cassettes are generated from single-stranded Ultramers or oligo pool

DNA using PCR with a reverse primer (**Figure 5.3B**). For these experiments, we assembled Libraries 1 and 2 from PCR-amplified mutant Ultramer cassettes and Library 3 from PCR-amplified mutant oligo pool cassettes (an example of this amplification for Library 1 is shown in **Figure 5.3C**). For comparison, we performed a fourth assembly with eBlock cassettes containing no sequence variation.

We performed PCR amplification of the oligo pools and Ultramers as well as the Golden Gate reactions using the base method described in our protocol, except for increasing input DNA from 40 to 200 fmol. We found similar transformation efficiencies using Ultramers and oligo pools, with both resulting in approximately 1.5×10^6 transformants, and a slightly higher efficiency with eBlocks, which resulted in approximately 7.0×10^6 transformants (**Figure 5.3D**). Thus, our protocol generates consistently high numbers of transformants with different types of input DNA, giving the user flexibility when designing their libraries.

To assess library quality, completed libraries were deep sequenced at a depth ranging from $5.9e^5$ to $2.9e^6$ reads (**Table D.5**). In contrast to other user-defined mutagenic protocols with high wild-type sequence carryover [82], no library contained more than 0.2% wild-type reads, and all libraries contained at least 96.7% of the desired variants (96.7, 99.9, 99.9%). The libraries ranged from 80.9-86.8% of on-target sequences. 9% of the oligo pool-derived library coded for chimera sequences that contained unintended combinations of designed mutations. These could have arisen during our deep sequencing preparation, as chimera formation is known to occur at low abundances during the associated PCR steps [3]. Alternatively, chimera formation could occur during the cassette generation step of our protocol; our analysis is unable to distinguish between these possibilities. We assessed library uniformity by comparing the theoretical frequency of different residues at each mutated position with the observed frequency seen in our deep sequencing data (**Figure 5.3E**). The relative frequencies observed varied between 0.34-1.52-fold as compared with expectations (n=62). Together, this data demonstrates that our Golden Gate protocol can generate user-defined combinatorial mutational libraries with almost complete coverage, little to no

wild-type carryover, and near-uniform individual mutational distributions.
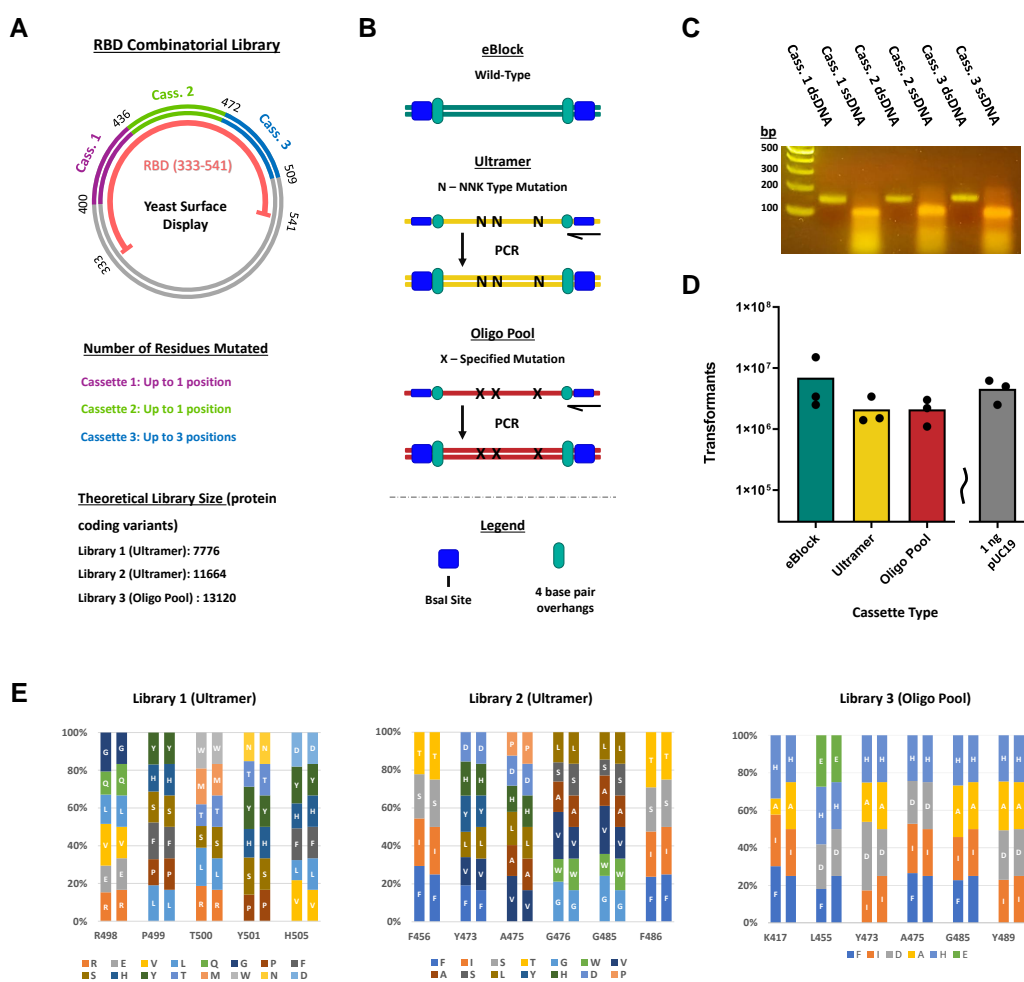
Figure 5.3: **Golden Gate assembly of SARS-CoV-2 S RBD deep mutational libraries.**
A. Schematic of the assembled plasmid for the S RBD combinatorial library. Unmutated S RBD residues 333-399 and 510-541 are encoded in the destination vector pIFU037 while residues 400-509 are encoded by three mutagenic cassettes. Cassettes can code for wild-type or mutant residues at each mutational site **B.** Different cassette DNA inputs for the Golden Gate assembly. eBlocks (dsDNA) are used as is, while Ultramers (mixed base-containing long oligonucleotides) and oligo pools are obtained as lyophilized ssDNA and require PCR synthesis using a reverse primer to generate dsDNA. **C.** Gel electrophoresis of three RBD library Ultramer cassettes before (ssDNA) and after PCR (dsDNA). Each cassette is 170 nts. **D.** Transformation efficiencies for Golden Gate reactions performed with pIFU037 and the three RBD library cassettes, each generated from different types of DNA. 25 μl reactions were performed in triplicate with 40 fmol of destination vector and cassettes. Numbers of transformants were calculated by plating serial dilutions of the recovered cells. pUC19 is shown as a control reaction for assessing efficiency of electrocompetent cells. **E.** Mutational distributions for RBD libraries. Expected (right bars) vs observed mutational frequency (left bars) at each mutated residue based on deep sequencing.

We also built and characterized a library encoding $1.1x10^7$ theoretical T7 RNAP to show that this technique is broadly applicable and that uniformity and coverage are maintained when our protocol is scaled up (data not shown).

5.3   Conclusions

Although the need for large, site-specific libraries for protein engineering workflows is almost universal, the methods for generating such libraries vary greatly. Modern DNA synthesis technology, which allows for the rapid creation of short DNA fragments containing user-defined mutations, has the potential to streamline these methods and make libraries more consistent and affordable. Here we provide a simple, broadly applicable protocol detailing how to build large libraries from synthetic DNA cassettes using Golden Gate assembly. To provide an accurate assessment of this protocol's capabilities, we benchmarked the obtainable transformation and incorporation efficiencies as a function of cassette number, reaction size, and input DNA. We then demonstrated its real-world applicability by creating libraries of different sizes in two orthogonal protein engineering systems. First, we generated libraries for the SARS-CoV-2 S RBD on the order of 10^4 members and demonstrated by deep sequencing that near-uniform, near-complete libraries can be made equally well using mixed base oligonucleotides or oligo pools. Subsequently, we built and characterized a library encoding 1.1×10^7 theoretical T7 RNAP to show that this technique is broadly applicable and that uniformity and coverage are maintained when our protocol is scaled up. We expect our benchmarking will facilitate implementation of cassette-based Golden Gate mutagenesis for the protein engineering and design community.

5.4   Methods

5.4.1   Construction of destination vectors and cassettes

Destination vectors were created by combining a source vector and a GFP-encoding insert from pYTK04710 or pEDA524 using Gibson Assembly [141] or restriction enzyme cloning. Oligo

pool and Ultramer derived double-stranded cassettes were obtained by performing PCR with single-stranded source DNA and a single reverse primer, followed by gel extraction. Sequences for primers, completed destination vectors, and wild-type versions of cassettes are listed in the Supporting Data online.

5.4.2 Performance and assessment of Golden Gate reactions

All Golden Gate reactions were performed using a base protocol with some small number of changes which are listed below. The base protocol is as follows. 40 fmol of a destination vector and 40 fmol of each cassette are combined in a PCR tube along with 20 units of BsaI HF-V2, 400 units of T4 DNA ligase, 2.5 μl of 10$\times$ T4 ligase buffer, and nuclease-free H_2O up to 25 μl. The reaction mixture is then cycled between a 37°C and 16°C PCR step for a total of 60 cycles, followed by a final 5 minute, 37°C step and a 10 minute, 65°C step. The resulting DNA is then cleaned and concentrated using an NEB Monarch PCR DNA Cleanup Kit and eluted in 6 μl of nuclease-free H_2O. All 6 μl of concentrated DNA are then transformed by electroporation into 50 μl of TransforMax EPI300 *E. coli* cells. Additionally, 0.1 ng of pUC19 in 6 μl of nuclease-free water was transformed into TransforMax cells to assess cell competency. Transformations were carried out in 1 mm electroporation cuvettes (BTX, Cat# 45-0124) with an electroporator (Eppendorf, Cat# 4309000027) set to 1200 V, which generated time constants between 4-6 ms. Immediately after electroporation, cells were resuspended in 1 ml of SOC media (SOB from BD Cat# 244310 with 20mM dextrose), incubated at 37°C for 1 hour, plated on LB agar containing the appropriate antibiotic, and were incubated overnight at 37°C. Numbers of transformants and percentages of GFP negative colonies were determined by plating serial dilutions of the transformed cells ranging from $10^2 - 10^8$ fold-diluted and counting the number of green and white colonies in the dilution that had the highest number of colonies totaling between 10 and 100. If no green colonies could be detected, the percentage incorporation was estimated as one over the total number of colonies in the lowest dilution, as extrapolated from the dilution containing all counted colonies. pUC19 transformation efficiencies were similarly determined by counting colonies from serial dilution plating, with the

70

resulting numbers being scaled to reflect the number of transformants per 1 ng of transformed pUC19. For libraries, the remaining transformants not used for dilutions were plated onto bioassay plates and incubated at 37°C. The next day colonies were scraped from the plates and mini prepped to obtain library DNA. For some experiments, the amount of input DNA or total reaction size was modified from the base protocol; these modifications are noted in the main text. Reaction size and input DNA along with the destination vector and cassettes used in each reaction are listed in **Table D.4**.

5.4.3   Characterization of Golden Gate assembled RBD constructs

Binding of GG and Wuhan-1 RBDs to ACE2/CC12.1 was assessed by yeast display as described by Francino-Urdaniz et al [3].

5.4.4   DNA deep sequencing

Deep sequencing prep was performed as described in the "Method B" protocol from Kowalsky et al [61]. In brief, we did two rounds of PCR with an ExoI clean up in between. In the first round we amplify the amplicon using customized primers for each specific sequence. These primers also contain the TruSeq illumina adapters. The second PCR inserts the TruSeq barcodes for sequencing and is performed exactly as described in the "Method B" protocol. The primers used for the deep sequencing preparation are given in Supporting Data online.

The first PCR thermocycler conditions for the RBD library are as follows:

PCR cycling conditions			
Steps	**Temperature**	**Time**	**Cycles**
Initial Denaturation	98C°C	60 sec	1
Denaturation	98°C	10 sec	18 cycles
Annealing	57°C	20 sec	
Extension	72°C	30 sec	
Final extension	72°C	10 min	1
Hold	4°C	infinity	

5.4.5 DNA deep sequencing analysis

For the analysis of the RBD library deep sequencing data, sequences were merged using an in-house merging code, essentially as previously described by Haas et al[121] The occurrence of mutations at each designed location were counted and summed. From these values, the corresponding distributions were generated (**Figure 5.3E**).

5.5 Acknowledgements

5.6 Author Contributions

Conceptualization: T.A.W., N.D., I.F.U., and Z.T.B. Experimental Design: T.A.W., N.D., I.F.U., and Z.T.B. Performed Experiments: N.D., I.F.U., and Z.T.B. Data Analysis: T.A.W., N.D.,

I.F.U., and Z.T.B. Wrote First Draft: N.D. Writing & Editing: T.A.W., N.D., I.F.U., and Z.T.B. Supervision: T.A.W. Funding Acquisition: T.A.W. and Z.T.B.

5.7   Data Availability

The plasmids pNDD003, pND004, and pND005 and their corresponding sequences have been deposited at AddGene (to be released upon publication). All raw deep sequencing reads have been deposited at the Sequencing Read Archive (SRA to be updated upon publication). Raw data used to prepare figures 5.1C, 5.2A-C, and 5.3D are present in Table D.4.

Chapter 6

Drifts and shifts in SARS-CoV-2 S RBD functional sequence space mapped by deep learning and deep mutational scanning

6.1 Introduction

An open question in biology is the allowable sequence variation of one side of a protein interface. This question has practical implications in the development of therapeutics and vaccine immunogens for mitigating infectious diseases, since mutations on surface glycoproteins of pathogens can perturb molecular recognition [142]. A well-known example relates to the COVID-19 pandemic. Owing to the extensive sequence diversity of SARS-CoV-2 subtypes at the relevant Spike RBD epitopes, only one of the seven regulatory approved monoclonal antibodies or cocktails is still currently effective (July 2023) (`https://www.cms.gov/monoclonal`).

Earlier efforts at assessing the potential SARS-CoV-2 RBD sequence diversity focused largely on the local mutational landscape (1-2 mutations away from the parental sequence). Starr et al. developed a yeast display platform for evaluating the sequence-function landscape of Wuhan Hu-1 S RBD by deep mutational scanning [65]. This landmark paper was followed by efforts from multiple groups at evaluating sequence tolerance for ACE2 recognition and/or propensity for escape from antibody-mediated virus neutralization [143, 144, 145, 3]. These studies were enormously helpful in predicting and prospectively identifying sequence variations on several early VOCs and in tracking the effectiveness of therapeutic antibodies. However, the emergence of Omicron that did not clearly descend from any of the previous VOCs [146, 147] was a siren call that a larger functional sequence

space is both available and accessible for this virus. This suggests that new methods are urgently needed to assess larger sequence functional diversity of viral glycoproteins.

Although various computational approaches have been described for assessing viral sequence variation at scale [148, 149], to our knowledge no experiments have been reported which assess large shifts in sequence variation of viral glycoproteins. Yet, the idea of experimentally assessing prospective sequence variation is not new in other fields. Neutral drift directed evolution is a well-established technique to further evolve existing enzymes under selective pressure to develop new functions [150, 151, 152]. These approaches have traditionally encoded sequence diversity using random mutagenesis. We hypothesized that, with the recent deep learning advances in protein design, we could design and measure larger shifts in protein sequence space by enriching mutational libraries with functional sequences.

In this paper we sought to prospectively identify the allowable sequence diversity of S RBD that maintains ACE2 recognition. We used computational design to develop several mutational libraries focused at the RBS, which we assessed for maintenance of ACE2 binding using yeast surface display and flow cytometry. Selected variants contain up to eighteen mutations away from the Omicron strain and show considerable plasticity of RBD sequence variation. We trained a predictive model to identify the prospective landscape of the ACE2-binding footprint on RBD, which was experimentally validated. Overall, our work identifies several discrete 'islands' in sequence space that delineate potential, and prospective, variation of the RBD Class 1 and Class 2 epitopes [50] which overlap with the RBS. Our work has broad implications for the design of next-generation immunogens for vaccines and escape mutation-resistant monoclonal antibody therapeutics.

6.2   Results

To ascertain the allowable sequence variation on RBD that still maintains potent binding to ACE2, we constructed several sequence-binding experimental datasets that represent large shifts in sequence space from an Omicron chimeric S RBD (SARS-CoV-2 S RBD (333-541) Wuhan-1 with

mutations K417N, S477N, E484A, Q498R, N501Y and Y505H). We denoted this Omicron chimeric S RBD wild-type or 'WT' as it was the starting point for all libraries. We designed and assembled five separate RBD mutational libraries covering the solvent-accessible positions of the Barnes and Bjorkman-defined Class 1 or Class 2 epitopes [50] around the RBS. Our first three libraries (LY003, LY005, LY006) used degenerate codons and a reduced genetic alphabet [153, 154]. Library LY005 has 7,776 variants containing up to 5 mutations per variant. Libraries LY003 consists of 39,585 with up to 7 mutations per library and LY006 consists of 20,736 variants with up to 6 mutations per variant. The other two libraries were designed by deep learning [155]: LY008 encodes 1,611 variants containing between 2 and 20 mutations with an average of 15 mutations per variant, and LY009 encodes 2,268 variants (range 9-16 mutations, mean 12). In sum, our combined libraries contain 71,976 variants with multiple mutations at 50 distinct positions in the Class I and/or Class II epitopes. We analyzed the sequence space covered by our libraries by PCA (**Fig 6.1B**). Compared with existing variants of concern (VOCs) and other previously described mutational libraries [65, 144], our libraries cover a larger and broader sequence space.

We used yeast surface display of S RBD variants to assess maintenance of binding to ACE2. Functional properties of yeast displayed S RBD variants correlate well with full-length S, as demonstrated by multiple groups (CITE). Our custom libraries were constructed from oligo pools using Golden Gate assembly [6] and transformed into yeast. After induction of S RBD surface expression, libraries were sorted at various ACE2 concentrations (range 1-1000 nM) using Fluorescence Activated Cell Sorting (FACS). At each concentration, we collected all cells maintaining binding to ACE2 (**Figure 6.1C**, **Figure C.19**) and deep sequenced the population along with a reference population. This experimental pipeline enabled a qualitative analysis of ACE2 binding, compared to WT, by correlating the probability of being sorted at each ACE2 concentration to three classification bins: "like WT", "worse than WT" and "no binding" (**Figure 6.1D**; see Methods).

We performed a number of experiments to assess the quality and consistency of the generated binding datasets. To test the internal consistency of the deep sequencing pipeline, we performed

76
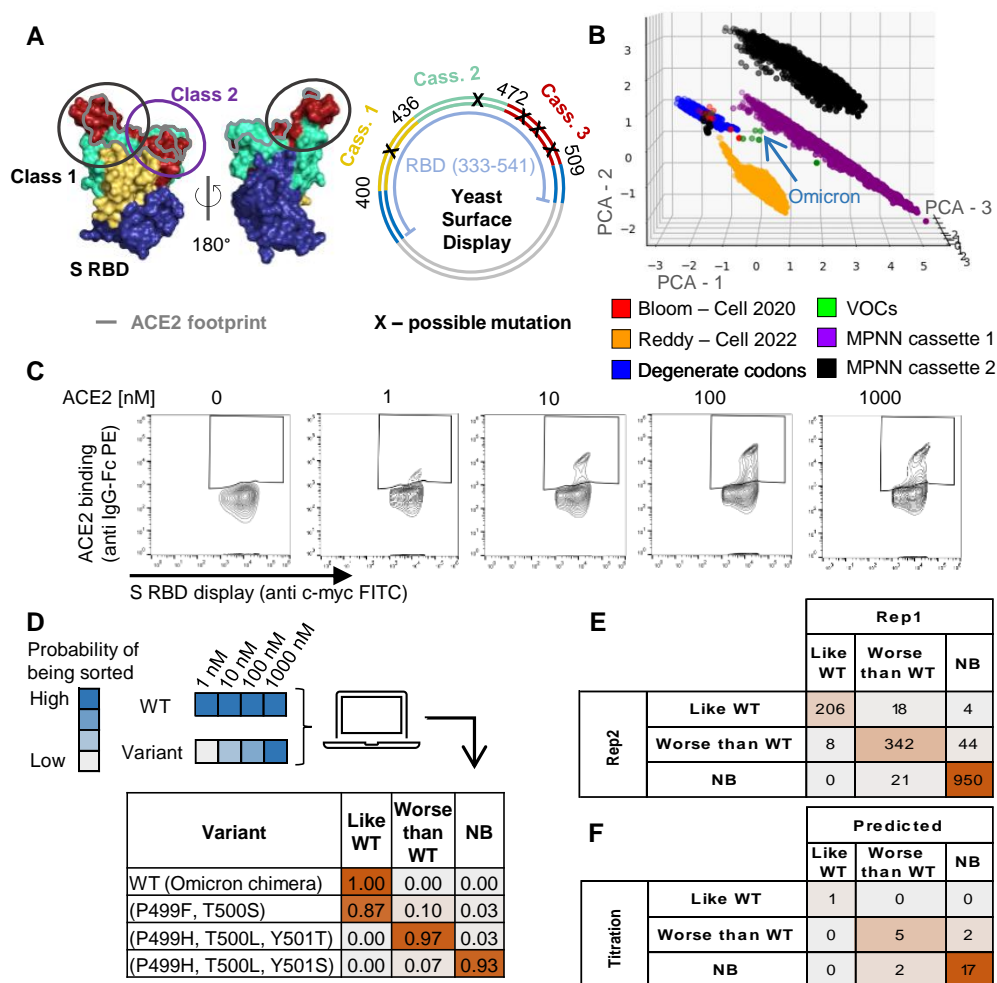
Figure 6.1: **Building S RBD libraries using yeast surface display.** **A.** R SBD structure showing the division of the ACE2 footprint in 3 cassettes (left) and yeast surface display plasmid (right). Cassette 1 covers residues 400 to 436 in yellow cassette 2 covers residues 437 to 472 in cyan, and Cassette 3 covers residues 473 to 509 in red. Positions that will not be mutated are shown in blue. **B.** Principal components analysis showing the sequence space covered by the designed libraries using a reduced library ("Degenerate codons" - blue) and Protein MPNN ("MPNN - Cassette 2" - black & "MPNN - Cassette 3" - purple) compared to the variants of concern ("VOCs" - green) and variants found in literature ("Bloom" [65] - red & "Reddy" [144] - orange). **C.** Collection gates for binding variants on ACE2 binding vs RBD display cytograms at 0nM to set the gate above noise, as well as collecting gates at 1nM, 10nM, 100nM and 1000nM ACE2. **D.** We analyze the probability of a variant of being sorted at each ACE2 concentration and compare it to the WT probability. By a computational analysis we can classify each variant into three classification bins: "like WT", "worse than WT" and "no binding". **E.** Biologial replicate comparison of the number of variants that fall into each classification bin. **F.** Comparison of the predicted classification bin of randomly selected variants to the binding observed on isogenic titration curves.

biological replicates on LY008 by separately transforming the plasmid library into yeast and sorting on different days (**Figure 6.1E**). We observed 93% agreement between replicates. Notably, only 0.25% (4/1593) of variants were classified as 'like WT' in one replicate and 'no binding' in the other replicate. Thus, the internal consistency of the deep sequencing pipeline is largely self-consistent. To test whether titrations of individual clones correlated with population measurements, we randomly selected 30 variants and performed isogenic titrations in replicate (**Figure C.20**). We classified variants binding worse than WT if they had at least a 10-fold decrease in KD relative to WT and at least a 2-fold increase in fluorescence over background at 1 micomolar ACE2. We observed 85% (22/26) agreement between isogenic titrations and population measurements (**Figure 6.1F**, SI Table X), showing reasonable agreement between measurements. To test for the possibility that RBD designs bind ACE2 non-specifically, we assessed the ability of lysozyme to bind the RBD library. While lysozyme binds non-specifically to cells, the fluorescence of the displaying population does not increase relative to the non-displaying cells indicating that it does not bind to the displayed RBD (**Figure C.21**). Overall, our library can assess ACE2 binding of large libraries.

To assess the structural variability of the RBD surface we looked at the location of the positions mutated of all the libraries compared to ACE2 footprint. Our libraries cover most of the Class 1 and Class 2 epitopes and only six of the 24 residues that comprise the RBS are not mutated in one of our libraries (**Figure 6.2A**). This demonstrates the large variability of the RBD sequence space of our study. Looking at each library individually, LY008 covers the surroundings of the RBS with only five of the 25 mutated positions belonging to the RBS. As it could be expected, the space covered by this library has a high tolerance for mutations and 13% of the variants bind like WT and 24% bind worse than WT. This has implications on loss of binding of therapeutics due to high likelihood of escape mutants. Opposite to LY008, 10 out of the 19 positions mutated in library LY009 belong to the RBS. Since the library has an average of 12 mutations per variant, multiple positions on the RBS are mutated in each variant. This many mutations are not tolerated since none of the variants in library LY009 bind to ACE2. In addition, 1929/2267 variants have the
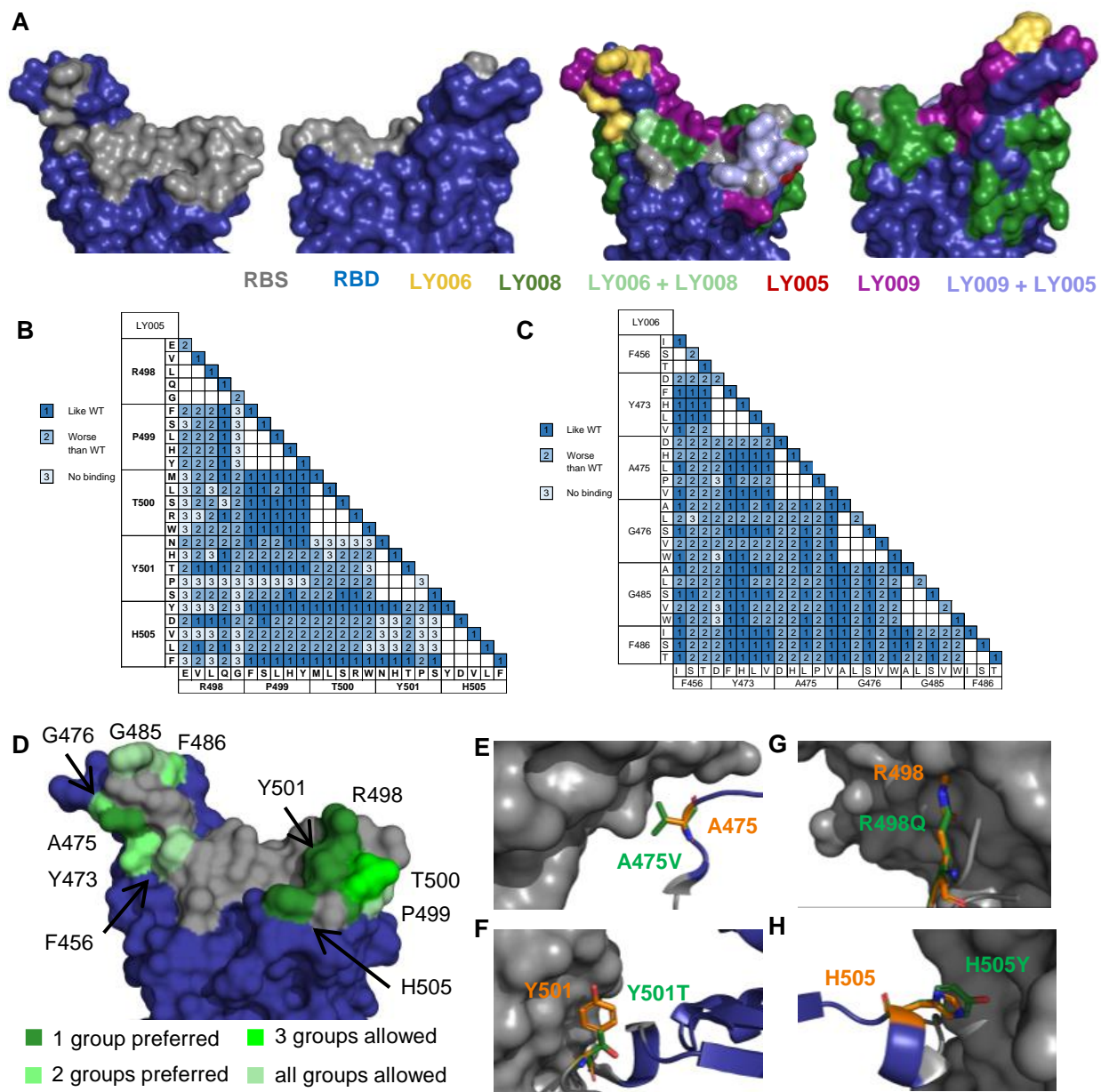
Figure 6.2: **Structural analysis of the mutational landscape. A.** Front and back of the RBD structure in blue with the RBS residues colored in grey (left). Allowable positions to mutate in each libray (right). **B - C.** Single and double mutants analysis on binding for library for (**B**) LY005 and (**C**) LY006. 1 - dark blue binds like WT, 2 - light blue binds worse than WT and 3 - light grey does not bind ACE2. **D.** RBD structure representing the tolerated mutations on Class 1 and Class 2 variants that maintain binding like WT. Different shades of gray show how many groups from the reduced amino acid library are seen. Unmutated residues are shown in dark blue and the ACE2 footprint in gray. **E - H.** Interaction between the RBD (dark blue cartoon) and ACE2 (gray surface) at positions with low tolerance to bind like WT. (**E**) A475V, (**F**) Y501T, (**G**) R498Q, (**H**) H505Y. Orange shows the WT residue structure and green the mutated amino acid.

combination of Y501N and H505N. Asparagine is in the same group as aspartic acid in the reduced alphabet used [154]. In the analysis of Class 2 epitope, we identify the combination of Y501N and H505D as deleterious (**Figure 6.2B**). We believe that some of the mutations in this library might be key to the interaction.

Library LY005 has up to 5 positions mutated to 5 different amino acids in the Class 1 epitope, 4 of these belong to the RBS. When analyzing single and double mutants (**Figure 6.2B**) we see that all of the single mutants except for Y501P maintain binding to ACE2. Moreover, Y501P is a deleterious mutation in any combination but the binding can be recovered when mutating the T500 residue to T500M/L/S/R/W. Although some combinations of mutations are widely tolerated such as P499 in combination with T500, 77% of the variants in the library completely lose binding to ACE2. This is in agreement with the complete loss of binding of library LY009 where multiple positions on the Class 2 epitope are mutated. On the other hand, 5% of the variants bind at least as good as WT.

Finally, library LY006 has 6 positions mutated to 3 or 5 different amino acids in the Class 2 epitope, 4 of these belong to the RBS. When analyzing single and double mutants (**Figure 6.2C**) we see that all of the single mutants maintain binding to ACE2. We see that the Class 2 epitope sequence is available for high sequence variability. Only five combinations of double mutants completely lose binding. When looking at the whole library, 42% of the variants maintain the ACE2 recognition (2% bind like WT and 40% bind worse than WT). We hypothesize that since the loop formed by residues 471-489 can change conformation between 'up' and 'down' positions [28] it is more flexible than the structure in class 2 allowing it to have a higher tolerance for mutations.

Looking closely at the variants in libraries LY005 and LY006 that maintain binding as good as WT, we analyzed what group of mutations from the reduced amino acid alphabet used [154] can be tolerated at each of the mutated positions. Although the combination of mutations has a large effect on binding [156], we noticed that some positions can be widely mutated while others

are more restricted (**Figure 6.2D**). In particular positions A475, R498, Y501 and H505 (**Figure 6.2E-H**) can only tolerate mutations from one group of amino acids. Mutation Y501T (**Figure 6.2F**) had not been previously observed in any of the variants of concern but it has appeared on the cryptic Ohaio variant identified on wastewater analysis by Mark Johnson lab (`https://twitter.com/SolidEvidence/status/1650304880160980992`). This is also the case of the reversion to the origins Wuhan-1 variant H505Y (**Figure 6.2H**).

6.3   Discussion

We developed a robust experimental tool to prospectively map protein variability. With it we have studied the tolerated sequence space on the RBD Class 1 and Class 2. From the structural analysis we have found that the Class 1 epitope is less tolerant than Class 2 epitope, as seen with libraries LY005 and LY009. This indicates that some mutations in this sequence space might be key to the interaction. However, we have also identified tolerated mutations in this region that have not been seen in existing variants, highlighting the benefit of prospectively mapping this interaction. Further analysis is needed to identify specific positions that have lower sequence tolerance but it is outside the reach of experimental methods. Therefore, we are training a computational model to learn from the experimental results and predict the effect on binding of a larger set of variants of RBD Class 1 and Class 2 epitopes.

Although the results are very promising, the developed tool has some limitations. (1) since the analysis is only qualitative and based on the binding of the WT, the dissociation constant of the WT has to be within the range of the variants in the library, and the WT must be in excess in the sorted library. (2) There is a maximum number of variants that can be sorted in a single library. We believe that this is mainly due to the library coverage obtained from the downstream sequencing. Our library LY003 had about 40k variants and we could not obtain clear results for the analysis. (3) The results are qualitative and associated with an error of the probability. In our hands we could only obtain 92% accuracy.

6.4   Methods

6.4.1   Design of mutagenic libraries

We designed a chimeric RBD with common mutations found on Omicron sub-variants (K417N, S477N, E484A, Q498R, N501Y and Y505H) that we define as WT. We have previously designed a plasmid containing the Wuhan-1 RBD gene [3] that we used as base plasmid with a few variations. To insert multiple customized mutations in each variant, we use the cassette-based mutagenesis approach previously described in Daffern et al. [6]. Therefore, to the base plasmid, we include the unmutated region of the RBD and BsaI sites with the right overhangs to insert the mutagenized cassettes.

We designed 6 libraries with different approaches. Libraries LY003, LY005 and LY006 were designed based on the structural analysis of the RBD-ACE2 interaction. To limit the number of mutations in these libraries, we used a reduced alphabet [154] that will maintain the protein structure. LY003 was made with custom oligo pools with mutations restricted to Class 2 epitopes. LY005 and LY006 were made with degenerate codons with mutations to Class 2 (LY005) or Class 3 (LY006) epitopes.

6.4.2   Preparation of the mutagenic libraries

Libraries were constructed exactly as described in Daffern et al [6] using degenerate oligonucleotides (LY005 and LY006) or custom oligo pools (LY003, LY008, LY009 and LY0010). Transformation and incorporation efficiencies for each library transformation are given in **Table D.7** and were also calculated as described.

The constructed libraries were midi prepped using ZymoPURE II plasmid midiprep kit to extract the plasmid. Yeast transformation was performed exactly as described [90] by transforming 5μg of DNA into chemically competent S. cerevisiae EBY100 [91]. Serial dilutions were plated on SDCAA agar plated and incubated for 3 days at 30°C to calculate the efficiency of the transforma-

tion (**Table D.7**). In parallel, the cells were grown in SDCAA (20g/L dextrose, 6.7g/L Difco yeast nitrogen base, 5g/L Bacto casamino acids, 5.4g/L Na_2HPO_4, and 8.56g/L $NaH_2PO_4 \cdot H_2O$) liquid media for 3 days at 30°C and 300rpm to saturation with 50 μg/mL Kanamycin and 1x PenStrep for 30h. Stocks were prepared at $OD_{600} = 1$ and at $OD_{600} = 10$ for screening and sorting respectively as in Whitehead et al. [94].

6.4.3   Screening and sorting libraries

For cell surface screening the cell stocks were grown in SDCAA at $OD_{600} = 0.5$ overnight at 30°C and 300rpm. The next day, the cells were induced in SGDCAA (18g/L galactose, 2g/L dextrose, 6.7g/L Difco yeast nitrogen base, 5g/L Bacto casamino acids, 5.4g/L Na_2HPO_4, and 8.56g/L $NaH_2PO_4 \cdot H_2O$) at $OD_{600} = 1$ at 22°C and 300 rpm for 22h with 50 μg/mL Kanamycin and 1× PenStrep for 30h. For sorting the libraries, 2% WT was spiked in on the SGDCAA inducing culture. The displaying cells are washed with PBSF (PBS containing 1g/l BSA) and stored on ice until used.

For WT titrations, 1×10^5 yeast cells were labeled with titrated concentrations of ACE2 ranging from 1pM to 1μM. Titrations were performed as described by Chao et al. [71] with an incubation time of 4h at 22°C and shaking. The cells were centrifuged and washed with 200μL PBSF. The cells were secondary labeled with 0.6 μL anti-c-myc-FITC (Miltenyi Biotec), 0.25 μL Goat anti-Human IgG Fc PE conjugate (Invitrogen Catalog # 12-4998-82) and 49.15μL PBSF for 10min at 4°C. The cells were washed and resuspended in 100μL PBSF. 25,000 cells were screened using a flow cytometer for each labeling concentration.

For sorting of the libraries, 3×10^7 cells were labeled at different concentrations of ACE2 in PBSF (1nM, 10nM, 100nM and 1000nM) for 1h at 22°C with shaking. The cells were centrifuged and washed with 1000μL PBSF. Each reaction was split into four Eppendorf tubes and individually labeled with 2.4 μL anti-c-myc-FITC (Miltenyi Biotec), 1 μL Goat anti-Human IgG Fc PE conjugate (Invitrogen Catalog # 12-4998-82) and 196.6 μL PBSF for 10 min at 4°C. Cells were

then centrifuged, washed with PBSF. The pellet was resuspended in 1000 μL and read on a flow cytometer to measure binding of the ACE2. The gates used for sorting were set up using cells not labeled with ACE2 and are shown in **Figure C.19**. The gates set are the following: an FSC/SSC$^+$ gate for isolation of yeast cells, FSC-H/FSC-A gate to discriminate single cells, a FSC-A/FITC$^+$ gate selects the cells displaying the RBD on their surface and from this last gate, everything above background noise by a PE$^+$/FITC$^+$ is collected. About 100-fold above library size cells were collected (1.0×10^6 for LY003, LY005 and LY006 and 5.0×10^5 for LY008 and LY009). A reference for each library was also collected as the cells that display in the FSC-A/FITC$^+$ gate. For the reference, about 500-fold above library size cells were collected (5.0×10^6 for LY003, LY005 and LY006 and 3.0×10^6 for LY008 and LY009). The collected cells were spun down and recovered in SDCAA with 50 μg/mL Kanamycin and 1$\times$ PenStrep for 30h. LY008 library was sorted in replicate in different days from different yeast stocks of the same Golden Gate reaction.

6.4.4   Isogenic variant titrations

Each library was separately plated on SDCAA Agar plates and grown at 30°C for 2 days. Individual colonies were selected and grown in SDCAA with 50 μg/mL Kanamycin and 1$\times$ PenStrep at 30°C and 300rpm for 30h. In parallel, specific variants were assembled using Golden Gate as previously described, and transformed into yeast. The cells were induced in SGDCAA with 50 μg/mL Kanamycin and 1$\times$ PenStrep at 22°C and 300rpm for 22h. Titrations were performed as previously described for the WT ranging from 1pM to 1μM.

6.4.5   Pseudovirus assay

Waiting on collaborators - will have results by the final submission

6.4.6   Polyspecificity assay

Lysozyme was biotinylated with EZ-link NHS-Biotin following manufacturer's instructions. 1×10^5 yeast cells were labeled with 250nM, 1μM and 5μM biotinylated lysozyme (Sigma L6876)

for 30 min at 22°C and shaking. The cells were centrifuged and washed with 200μL PBSF. The cells were secondary labeled with 0.6 μL anti-c-myc-FITC (Miltenyi Biotec), 0.25 μL SAPE (ThermoFisher S688) and 49.15μL PBSF for 10min at 4°C. The cells were washed and resuspended in 100μL PBSF. 25,000 cells were screened using a flow cytometer for each labeling concentration.

6.4.7   Deep sequencing preparation

The DNA was prepared for deep sequencing following the "Method B" protocol from Kowalsky et al [61]. The amplicon was amplified as described in Daffern et al. [6] Samples were then further purified using Agencourt Ampure XP beads (Beckman Coulter), quantified using PicoGreen (ThermoFisher), pooled, and sequenced on an Illumina MiSeq using 2×250 bp paired-end reads at Rush Genomics and Microbiome Core Facility (Rush University, Chicago, IL).

Chapter 7

Conclusions and future work

7.1   Conclusions

I have built an easy to use, quick and robust tool to identify escape mutants. With it, I identified 97 S RBD mutants that can escape recognition by at least one of the antibodies tested. Of these 40/54 (74%) of the single nucleotide mutations had not been previously identified in literature. Additionally, I identified >70% of the mutations on the RBD Class 1 and Class 2 epitopes that emerged with the Omicron variants. With the identified mutations of escape for each of the antibodies in the study, I informed an engineering campaign to increase the affinity of an antibody towards the RBD and develop broadly neutralizing antibodies [5]. The original antibodies were isolated from SARS-CoV-2 Wuhan-1 convalescent patients and were not able to neutralize Delta or Omicron variants. After engineering, these antibodies could neutralize both variants in addition of Wuhan-1 and Beta.

I have built highly mutational libraries by benchmarking a cassette-based method using Golden Gate. I believe that this method will be used by many researchers to make library generation easier and faster. I have used this method to study the mutational sequence space on the RBD Class 1 and Class 2 epitope and its effects on binding. I can predict with a 92% accuracy wich is the probability of a given mutation to bind like WT, worse than WT or lose binding to ACE2. Although there is still some work to be done to expand the platform, I have shown that experimentally, the

residues with lower tolerance to mutations can be identified. With the experimental results we are training a computational model to extrapolate the results to a wider set of variants and with it we will identify the lower mutational space and therefore the potential target of therapies. While doing so, I have seen variants that bind like WT with mutations that have not been seen in existing variants of concern. Therefore, prospectively mapping the available sequence variability is not only beneficial to identify potential targets for therapeutics but also to find the allowed mutations in a specific position. I think that this tool will allow us to have a better understanding of new targets for antibody therapies and vaccine immunogens.

Overall, I have a good understanding of the antigenic variability of the RBD, and I have built robust tools on the way. I have also proven that deep mutational scanning with next generation sequencing is an effective method to study protein-protein interactions and can be used to inform the engineering of therapeutics and vaccine immunogens. All the tools were developed using SARS-CoV-2 S RBD as my system, but they are not specific for it. I believe that they can be widely applied to understand other protein-protein interactions and can inform further engineering efforts.

7.2   Future work

Future directions of this work could include using the knowledge of the mutational space on RBM to engineer broadly neutralizing antibodies or small molecules to bind the low variability region. The aims would be: 1. To identify the antibody germline that binds to the identified region[50] and choose a starting antibody. Design a library for the antibody, mutating the residues that interact directly with the RBD, as well as the surroundings to allow changes in flexibility of the loops. Finally, display the library on the surface of yeast. 2. Express the variant of RBD to engineer the antibody against using insect cells. 3. Perform several rounds of selection similar to Zhao et al.[5]. 4. To confirm that the engineered antibody is broadly neutralizing, the antibody should be expressed and scanned for escape mutants using an RBD library.

Another direction in which this work could be continued is to expand these tools to different

proteins. This work has focused on a binding protein such that there is a direct correlation between antibody binding and neutralization. It would be interesting to translate the developed tools to a non-binding protein, for example the NTD in the spike protein. The NTD, as well as the RBS, is targeted by potent neutralizing antibodies such as 4A8 [49]. Although this would have an additional challenge to assess functionality of the designed variants since its exact function still remains unknown[8]. We attempted to pursue this idea using the Influenza Neuraminidase (NA) surface protein whose function and structure is known[157] and neutralizing antibodies have been isolated[158]. The biggest challenge was to maintain enzymatic activity while displaying the enzyme on the surface of yeast. We designed a platform to map the mutational space of NA that would not affect the active site. However, we could not get consistent proof of NA activity of the platform.

In addition, the tools that we have built will be used across different new projects by the next generations of graduate students. An example is a recently NIH funded project to redesign the head of the Influenza hemagglutinin.

Bibliography

[1] I. M. Francino-Urdaniz and T. A. Whitehead, "An overview of methods for the structural and functional mapping of epitopes recognized by anti-SARS-CoV-2 antibodies," RSC Chem Biol, vol. 2, pp. 1580–1589, Dec. 2021.

[2] B. B. Banach, G. Cerutti, A. S. Fahad, C.-H. Shen, M. Oliveira De Souza, P. S. Katsamba, Y. Tsybovsky, P. Wang, M. S. Nair, Y. Huang, I. M. Francino-Urdániz, P. J. Steiner, M. Gutiérrez-González, L. Liu, S. N. López Acevedo, A. F. Nazzari, J. R. Wolfe, Y. Luo, A. S. Olia, I.-T. Teng, J. Yu, T. Zhou, E. R. Reddem, J. Bimela, X. Pan, B. Madan, A. D. Laflin, R. Nimrania, K.-Y. Yuen, T. A. Whitehead, D. D. Ho, P. D. Kwong, L. Shapiro, and B. J. DeKosky, "Paired heavy- and light-chain signatures contribute to potent SARS-CoV-2 neutralization in public antibody responses," Cell Rep., vol. 37, p. 109771, Oct. 2021.

[3] I. M. Francino-Urdaniz, P. J. Steiner, M. B. Kirby, F. Zhao, C. M. Haas, S. Barman, E. R. Rhodes, A. C. Leonard, L. Peng, K. G. Sprenger, J. G. Jardine, and T. A. Whitehead, "One-shot identification of SARS-CoV-2 S RBD escape mutants using yeast screening," Cell Rep., vol. 36, p. 109627, Aug. 2021.

[4] T. F. Rogers, F. Zhao, D. Huang, N. Beutler, A. Burns, W.-T. He, O. Limbo, C. Smith, G. Song, J. Woehl, L. Yang, R. K. Abbott, S. Callaghan, E. Garcia, J. Hurtado, M. Parren, L. Peng, S. Ramirez, J. Ricketts, M. J. Ricciardi, S. A. Rawlings, N. C. Wu, M. Yuan, D. M. Smith, D. Nemazee, J. R. Teijaro, J. E. Voss, I. A. Wilson, R. Andrabi, B. Briney, E. Landais, D. Sok, J. G. Jardine, and D. R. Burton, "Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model," Science, vol. 369, pp. 956–963, Aug. 2020.

[5] F. Zhao, C. Keating, G. Ozorowski, N. Shaabani, I. M. Francino-Urdaniz, S. Barman, O. Limbo, A. Burns, P. Zhou, M. J. Ricciardi, J. Woehl, Q. Tran, H. L. Turner, L. Peng, D. Huang, D. Nemazee, R. Andrabi, D. Sok, J. R. Teijaro, T. A. Whitehead, A. B. Ward, D. R. Burton, and J. G. Jardine, "Engineering SARS-CoV-2 neutralizing antibodies for increased potency and reduced viral escape pathways," iScience, vol. 25, p. 104914, Sept. 2022.

[6] N. Daffern, I. Francino-Urdaniz, Z. T. Baumer, and T. A. Whitehead, "Benchmarking cassette-based deep mutagenesis by golden gate assembly." Apr. 2023.

[7] G. M. Cherf and J. R. Cochran, "Applications of yeast surface display for protein engineering," Methods Mol. Biol., vol. 1319, pp. 155–175, 2015.

[8] C. B. Jackson, M. Farzan, B. Chen, and H. Choe, "Mechanisms of SARS-CoV-2 entry into cells," Nat. Rev. Mol. Cell Biol., vol. 23, pp. 3–20, Jan. 2022.

[9] X. Peng, J. Wang, W. Peng, F.-X. Wu, and Y. Pan, "Protein–protein interactions: detection, reliability assessment and applications," Brief. Bioinform., vol. 18, pp. 798–819, July 2016.

[10] D. W. Kulp and W. R. Schief, "Advances in structure-based vaccine design," Curr. Opin. Virol., vol. 3, pp. 322–331, June 2013.

[11] M. C. Crank, T. J. Ruckwardt, M. Chen, K. M. Morabito, E. Phung, P. J. Costner, L. A. Holman, S. P. Hickman, N. M. Berkowitz, I. J. Gordon, G. V. Yamshchikov, M. R. Gaudinski, A. Kumar, L. A. Chang, S. M. Moin, J. P. Hill, A. T. DiPiazza, R. M. Schwartz, L. Kueltzo, J. W. Cooper, P. Chen, J. A. Stein, K. Carlton, J. G. Gall, M. C. Nason, P. D. Kwong, G. L. Chen, J. R. Mascola, J. S. McLellan, J. E. Ledgerwood, B. S. Graham, and VRC 317 Study Team, "A proof of concept for structure-based vaccine design targeting RSV in humans," Science, vol. 365, pp. 505–509, Aug. 2019.

[12] D. C. Ekiert, G. Bhabha, M.-A. Elsliger, R. H. E. Friesen, M. Jongeneelen, M. Throsby, J. Goudsmit, and I. A. Wilson, "Antibody recognition of a highly conserved influenza virus epitope," Science, vol. 324, pp. 246–251, Apr. 2009.

[13] M. Throsby, E. van den Brink, M. Jongeneelen, L. L. M. Poon, P. Alard, L. Cornelissen, A. Bakker, F. Cox, E. van Deventer, Y. Guan, J. Cinatl, J. ter Meulen, I. Lasters, R. Carsetti, M. Peiris, J. de Kruif, and J. Goudsmit, "Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells," PLoS One, vol. 3, p. e3942, Dec. 2008.

[14] J. Sui, W. C. Hwang, S. Perez, G. Wei, D. Aird, L.-M. Chen, E. Santelli, B. Stec, G. Cadwell, M. Ali, H. Wan, A. Murakami, A. Yammanuru, T. Han, N. J. Cox, L. A. Bankston, R. O. Donis, R. C. Liddington, and W. A. Marasco, "Structural and functional bases for broad-spectrum neutralization of avian and human influenza a viruses," Nat. Struct. Mol. Biol., vol. 16, pp. 265–273, Mar. 2009.

[15] R. Nachbagauer, J. Feser, A. Naficy, D. I. Bernstein, J. Guptill, E. B. Walter, F. Berlanda-Scorza, D. Stadlbauer, P. C. Wilson, T. Aydillo, M. A. Behzadi, D. Bhavsar, C. Bliss, C. Capuano, J. M. Carreño, V. Chromikova, C. Claeys, L. Coughlan, A. W. Freyn, C. Gast, A. Javier, K. Jiang, C. Mariottini, M. McMahon, M. McNeal, A. Solórzano, S. Strohmeier, W. Sun, M. Van der Wielen, B. L. Innis, A. García-Sastre, P. Palese, and F. Krammer, "A chimeric hemagglutinin-based universal influenza virus vaccine approach induces broad and long-lasting immunity in a randomized, placebo-controlled phase I trial," Nat. Med., vol. 27, pp. 106–114, Jan. 2021.

[16] E. Malito, A. Carfi, and M. J. Bottomley, "Protein crystallography in vaccine research and development," Int. J. Mol. Sci., vol. 16, pp. 13106–13140, June 2015.

[17] G. E. Morris, "Epitope mapping," Methods Mol. Biol., vol. 295, pp. 255–268, 2005.

[18] J.-P. Renaud, A. Chari, C. Ciferri, W.-T. Liu, H.-W. Rémigy, H. Stark, and C. Wiesmann, "Cryo-EM in drug discovery: achievements, limitations and prospects," Nat. Rev. Drug Discov., vol. 17, pp. 471–492, July 2018.

[19] C. Katz, L. Levy-Beladev, S. Rotem-Bamberger, T. Rito, S. G. D. Rüdiger, and A. Friedler, "Studying protein-protein interactions using peptide arrays," Chem. Soc. Rev., vol. 40, pp. 2131–2145, May 2011.

[20] R. Liu, A. M. Enstrom, and K. S. Lam, "Combinatorial peptide library methods for immuno-biology research," Exp. Hematol., vol. 31, pp. 11–30, Jan. 2003.

[21] J. Pande, M. M. Szewczyk, and A. K. Grover, "Phage display: concept, innovations, applications and future," Biotechnol. Adv., vol. 28, pp. 849–858, July 2010.

[22] M. B. Irving, O. Pan, and J. K. Scott, "Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics," Curr. Opin. Chem. Biol., vol. 5, pp. 314–324, June 2001.

[23] H. Sun, L. Ma, L. Wang, P. Xiao, H. Li, M. Zhou, and D. Song, "Research advances in hydrogen-deuterium exchange mass spectrometry for protein epitope mapping," Anal. Bioanal. Chem., vol. 413, pp. 2345–2359, Apr. 2021.

[24] D. M. Fowler and S. Fields, "Deep mutational scanning: a new style of protein science," Nat. Methods, vol. 11, pp. 801–807, Aug. 2014.

[25] K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, and R. F. Garry, "The proximal origin of SARS-CoV-2," Nat. Med., vol. 26, pp. 450–452, Apr. 2020.

[26] H. Else, "How a torrent of COVID science changed research publishing - in seven charts," Nature, vol. 588, p. 553, Dec. 2020.

[27] T. Noy-Porat, E. Makdasi, R. Alcalay, A. Mechaly, Y. Levy, A. Bercovich-Kinori, A. Zauberman, H. Tamir, Y. Yahalom-Ronen, M. Israeli, E. Epstein, H. Achdout, S. Melamed, T. Chitlaru, S. Weiss, E. Peretz, O. Rosen, N. Paran, S. Yitzhaki, S. C. Shapira, T. Israely, O. Mazor, and R. Rosenfeld, "A panel of human neutralizing mabs targeting SARS-CoV-2 spike at multiple epitopes," Nat. Commun., vol. 11, p. 4303, Aug. 2020.

[28] M. Yuan, N. C. Wu, X. Zhu, C.-C. D. Lee, R. T. Y. So, H. Lv, C. K. P. Mok, and I. A. Wilson, "A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV," Science, vol. 368, pp. 630–633, May 2020.

[29] X. Zeng, L. Li, J. Lin, X. Li, B. Liu, Y. Kong, S. Zeng, J. Du, H. Xiao, T. Zhang, S. Zhang, and J. Liu, "Isolation of a human monoclonal antibody specific for the receptor binding domain of SARS-CoV-2 using a competitive phage biopanning strategy," Antib Ther, vol. 3, pp. 95–100, Apr. 2020.

[30] A. Baum, B. O. Fulton, E. Wloga, R. Copin, K. E. Pascal, V. Russo, S. Giordano, K. Lanza, N. Negron, M. Ni, Y. Wei, G. S. Atwal, A. J. Murphy, N. Stahl, G. D. Yancopoulos, and

C. A. Kyratsous, "Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies," Science, vol. 369, pp. 1014–1018, Aug. 2020.

[31] A. J. Greaney, T. N. Starr, P. Gilchuk, S. J. Zost, E. Binshtein, A. N. Loes, S. K. Hilton, J. Huddleston, R. Eguia, K. H. D. Crawford, A. S. Dingens, R. S. Nargi, R. E. Sutton, N. Suryadevara, P. W. Rothlauf, Z. Liu, S. P. J. Whelan, R. H. Carnahan, J. E. Crowe, Jr, and J. D. Bloom, "Complete mapping of mutations to the SARS-CoV-2 spike Receptor-Binding domain that escape antibody recognition," Cell Host Microbe, vol. 29, pp. 44–57.e9, Jan. 2021.

[32] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, "Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein," Cell, vol. 183, p. 1735, Dec. 2020.

[33] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, and Q. Zhou, "Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2," Science, vol. 367, pp. 1444–1448, Mar. 2020.

[34] M. A. Tortorici and D. Veesler, "Structural insights into coronavirus entry," Adv. Virus Res., vol. 105, pp. 93–116, Aug. 2019.

[35] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan, "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation," bioRxiv, Feb. 2020.

[36] D. Pinto, Y.-J. Park, M. Beltramello, A. C. Walls, M. A. Tortorici, S. Bianchi, S. Jaconi, K. Culap, F. Zatta, A. De Marco, A. Peter, B. Guarino, R. Spreafico, E. Cameroni, J. B. Case, R. E. Chen, C. Havenar-Daughton, G. Snell, A. Telenti, H. W. Virgin, A. Lanzavecchia, M. S. Diamond, K. Fink, D. Veesler, and D. Corti, "Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody," Nature, vol. 583, pp. 290–295, July 2020.

[37] M. Yuan, H. Liu, N. C. Wu, C.-C. D. Lee, X. Zhu, F. Zhao, D. Huang, W. Yu, Y. Hua, H. Tien, T. F. Rogers, E. Landais, D. Sok, J. G. Jardine, D. R. Burton, and I. A. Wilson, "Structural basis of a shared antibody response to SARS-CoV-2," Science, vol. 369, pp. 1119–1123, Aug. 2020.

[38] Y. Wu, F. Wang, C. Shen, W. Peng, D. Li, C. Zhao, Z. Li, S. Li, Y. Bi, Y. Yang, Y. Gong, H. Xiao, Z. Fan, S. Tan, G. Wu, W. Tan, X. Lu, C. Fan, Q. Wang, Y. Liu, C. Zhang, J. Qi, G. F. Gao, F. Gao, and L. Liu, "A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2," Science, vol. 368, pp. 1274–1278, June 2020.

[39] M. McCallum, A. De Marco, F. A. Lempp, M. A. Tortorici, D. Pinto, A. C. Walls, M. Beltramello, A. Chen, Z. Liu, F. Zatta, S. Zepeda, J. di Iulio, J. E. Bowen, M. Montiel-Ruiz, J. Zhou, L. E. Rosen, S. Bianchi, B. Guarino, C. S. Fregni, R. Abdelnabi, S.-Y. C. Foo, P. W. Rothlauf, L.-M. Bloyet, F. Benigni, E. Cameroni, J. Neyts, A. Riva, G. Snell, A. Telenti, S. P. J. Whelan, H. W. Virgin, D. Corti, M. S. Pizzuto, and D. Veesler, "N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2," Cell, vol. 184, pp. 2332–2347.e16, Apr. 2021.

[40] G. Cerutti, Y. Guo, T. Zhou, J. Gorman, M. Lee, M. Rapp, E. R. Reddem, J. Yu, F. Bahna, J. Bimela, Y. Huang, P. S. Katsamba, L. Liu, M. S. Nair, R. Rawi, A. S. Olia, P. Wang, B. Zhang, G.-Y. Chuang, D. D. Ho, Z. Sheng, P. D. Kwong, and L. Shapiro, "Potent SARS-CoV-2 neutralizing antibodies directed against spike n-terminal domain target a single supersite," Cell Host Microbe, vol. 29, pp. 819–833.e7, May 2021.

[41] X. Chi, R. Yan, J. Zhang, G. Zhang, Y. Zhang, M. Hao, Z. Zhang, P. Fan, Y. Dong, Y. Yang, Z. Chen, Y. Guo, J. Zhang, Y. Li, X. Song, Y. Chen, L. Xia, L. Fu, L. Hou, J. Xu, C. Yu, J. Li, Q. Zhou, and W. Chen, "A neutralizing human antibody binds to the n-terminal domain of the spike protein of SARS-CoV-2," Science, vol. 369, pp. 650–655, Aug. 2020.

[42] B. Ju, Q. Zhang, J. Ge, R. Wang, J. Sun, X. Ge, J. Yu, S. Shan, B. Zhou, S. Song, X. Tang, J. Yu, J. Lan, J. Yuan, H. Wang, J. Zhao, S. Zhang, Y. Wang, X. Shi, L. Liu, J. Zhao, X. Wang, Z. Zhang, and L. Zhang, "Human neutralizing antibodies elicited by SARS-CoV-2 infection," Nature, vol. 584, pp. 115–119, Aug. 2020.

[43] Y. Guo, L. Huang, G. Zhang, Y. Yao, H. Zhou, S. Shen, B. Shen, B. Li, X. Li, Q. Zhang, M. Chen, D. Chen, J. Wu, D. Fu, X. Zeng, M. Feng, C. Pi, Y. Wang, X. Zhou, M. Lu, Y. Li, Y. Fang, Y.-Y. Lu, X. Hu, S. Wang, W. Zhang, G. Gao, F. Adrian, Q. Wang, F. Yu, Y. Peng, A. G. Gabibov, J. Min, Y. Wang, H. Huang, A. Stepanov, W. Zhang, Y. Cai, J. Liu, Z. Yuan, C. Zhang, Z. Lou, F. Deng, H. Zhang, C. Shan, L. Schweizer, K. Sun, and Z. Rao, "A SARS-CoV-2 neutralizing antibody with extensive spike binding coverage and modified for optimal therapeutic outcomes," Nat. Commun., vol. 12, p. 2623, May 2021.

[44] R. Shi, C. Shan, X. Duan, Z. Chen, P. Liu, J. Song, T. Song, X. Bi, C. Han, L. Wu, G. Gao, X. Hu, Y. Zhang, Z. Tong, W. Huang, W. J. Liu, G. Wu, B. Zhang, L. Wang, J. Qi, H. Feng, F.-S. Wang, Q. Wang, G. F. Gao, Z. Yuan, and J. Yan, "A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2," Nature, vol. 584, pp. 120–124, Aug. 2020.

[45] D. Fu, G. Zhang, Y. Wang, Z. Zhang, H. Hu, S. Shen, J. Wu, B. Li, X. Li, Y. Fang, J. Liu, Q. Wang, Y. Zhou, W. Wang, Y. Li, Z. Lu, X. Wang, C. Nie, Y. Tian, D. Chen, Y. Wang, X. Zhou, Q. Wang, F. Yu, C. Zhang, C. Deng, L. Zhou, G. Guan, N. Shao, Z. Lou, F. Deng, H. Zhang, X. Chen, M. Wang, L. Liu, Z. Rao, and Y. Guo, "Structural basis for SARS-CoV-2 neutralizing antibodies with novel binding epitopes," PLoS Biol., vol. 19, p. e3001209, May 2021.

[46] C. O. Barnes, A. P. West, Jr, K. E. Huey-Tubman, M. A. G. Hoffmann, N. G. Sharaf, P. R. Hoffman, N. Koranda, H. B. Gristick, C. Gaebler, F. Muecksch, J. C. C. Lorenzi, S. Finkin, T. Hägglöf, A. Hurley, K. G. Millard, Y. Weisblum, F. Schmidt, T. Hatziioannou, P. D. Bieniasz, M. Caskey, D. F. Robbiani, M. C. Nussenzweig, and P. J. Bjorkman, "Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies," Cell, vol. 182, pp. 828–842.e16, Aug. 2020.

[47] C. G. Rappazzo, L. V. Tse, C. I. Kaku, D. Wrapp, M. Sakharkar, D. Huang, L. M. Deveau, T. J. Yockachonis, A. S. Herbert, M. B. Battles, C. M. O'Brien, M. E. Brown, J. C. Geoghegan, J. Belk, L. Peng, L. Yang, Y. Hou, T. D. Scobey, D. R. Burton, D. Nemazee, J. M. Dye, J. E. Voss, B. M. Gunn, J. S. McLellan, R. S. Baric, L. E. Gralinski, and L. M. Walker,

"Broad and potent activity against SARS-like viruses by an engineered human monoclonal antibody," Science, vol. 371, pp. 823–829, Feb. 2021.

[48] Z. Lv, Y.-Q. Deng, Q. Ye, L. Cao, C.-Y. Sun, C. Fan, W. Huang, S. Sun, Y. Sun, L. Zhu, Q. Chen, N. Wang, J. Nie, Z. Cui, D. Zhu, N. Shaw, X.-F. Li, Q. Li, L. Xie, Y. Wang, Z. Rao, C.-F. Qin, and X. Wang, "Structural basis for neutralization of SARS-CoV-2 and SARS-CoV by a potent therapeutic antibody," Science, vol. 369, pp. 1505–1509, Sept. 2020.

[49] L. Liu, P. Wang, M. S. Nair, J. Yu, M. Rapp, Q. Wang, Y. Luo, J. F.-W. Chan, V. Sahi, A. Figueroa, X. V. Guo, G. Cerutti, J. Bimela, J. Gorman, T. Zhou, Z. Chen, K.-Y. Yuen, P. D. Kwong, J. G. Sodroski, M. T. Yin, Z. Sheng, Y. Huang, L. Shapiro, and D. D. Ho, "Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike," Nature, vol. 584, pp. 450–456, Aug. 2020.

[50] C. O. Barnes, C. A. Jette, M. E. Abernathy, K.-M. A. Dam, S. R. Esswein, H. B. Gristick, A. G. Malyutin, N. G. Sharaf, K. E. Huey-Tubman, Y. E. Lee, D. F. Robbiani, M. C. Nussenzweig, A. P. West, Jr, and P. J. Bjorkman, "SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies," Nature, vol. 588, pp. 682–687, Dec. 2020.

[51] N. Mishra, X. Huang, S. Joshi, C. Guo, J. Ng, R. Thakkar, Y. Wu, X. Dong, Q. Li, R. S. Pinapati, E. Sullivan, A. Caciula, R. Tokarz, T. Briese, J. Lu, and W. I. Lipkin, "Immunoreactive peptide maps of SARS-CoV-2," Commun Biol, vol. 4, p. 225, Feb. 2021.

[52] Y. Li, M.-L. Ma, Q. Lei, F. Wang, W. Hong, D.-Y. Lai, H. Hou, Z.-W. Xu, B. Zhang, H. Chen, C. Yu, J.-B. Xue, Y.-X. Zheng, X.-N. Wang, H.-W. Jiang, H.-N. Zhang, H. Qi, S.-J. Guo, Y. Zhang, X. Lin, Z. Yao, J. Wu, H. Sheng, Y. Zhang, H. Wei, Z. Sun, X. Fan, and S.-C. Tao, "Linear epitope landscape of the SARS-CoV-2 spike protein constructed from 1,051 COVID-19 patients," Cell Rep., vol. 34, p. 108915, Mar. 2021.

[53] L. Farrera-Soler, J.-P. Daguer, S. Barluenga, O. Vadas, P. Cohen, S. Pagano, S. Yerly, L. Kaiser, N. Vuilleumier, and N. Winssinger, "Identification of immunodominant linear epitopes from SARS-CoV-2 patient plasma," PLoS One, vol. 15, p. e0238089, Sept. 2020.

[54] S. Ravichandran, E. M. Coyle, L. Klenow, J. Tang, G. Grubbs, S. Liu, T. Wang, H. Golding, and S. Khurana, "Antibody signature induced by SARS-CoV-2 spike protein immunogens in rabbits," Sci. Transl. Med., vol. 12, July 2020.

[55] Y. Li, D.-Y. Lai, Q. Lei, Z.-W. Xu, F. Wang, H. Hou, L. Chen, J. Wu, Y. Ren, M.-L. Ma, B. Zhang, H. Chen, C. Yu, J.-B. Xue, Y.-X. Zheng, X.-N. Wang, H.-W. Jiang, H.-N. Zhang, H. Qi, S.-J. Guo, Y. Zhang, X. Lin, Z. Yao, P. Pang, D. Shi, W. Wang, X. Yang, J. Zhou, H. Sheng, Z. Sun, H. Shan, X. Fan, and S.-C. Tao, "Systematic evaluation of IgG responses to SARS-CoV-2 spike protein-derived peptides for monitoring COVID-19 patients," Cell. Mol. Immunol., vol. 18, pp. 621–631, Mar. 2021.

[56] C. M. Poh, G. Carissimo, B. Wang, S. N. Amrun, C. Y.-P. Lee, R. S.-L. Chee, S.-W. Fong, N. K.-W. Yeo, W.-H. Lee, A. Torres-Ruesta, Y.-S. Leo, M. I.-C. Chen, S.-Y. Tan, L. Y. A. Chai, S. Kalimuddin, S. S. G. Kheng, S.-Y. Thien, B. E. Young, D. C. Lye, B. J. Hanson, C.-I. Wang, L. Renia, and L. F. P. Ng, "Two linear epitopes on the SARS-CoV-2 spike protein

that elicit neutralising antibodies in COVID-19 patients," Nat. Commun., vol. 11, p. 2806, June 2020.

[57] P. V. Raghuvamsi, N. K. Tulsian, F. Samsudin, X. Qian, K. Purushotorman, G. Yue, M. M. Kozma, W. Y. Hwa, J. Lescar, P. J. Bond, P. A. MacAry, and G. S. Anand, "SARS-CoV-2 S protein:ACE2 interaction reveals novel allosteric targets," Elife, vol. 10, Feb. 2021.

[58] B. E. Jones, P. L. Brown-Augsburger, K. S. Corbett, K. Westendorf, J. Davies, T. P. Cujec, C. M. Wiethoff, J. L. Blackbourne, B. A. Heinz, D. Foster, R. E. Higgs, D. Balasubramaniam, L. Wang, Y. Zhang, E. S. Yang, R. Bidshahri, L. Kraft, Y. Hwang, S. Žentelis, K. R. Jepson, R. Goya, M. A. Smith, D. W. Collins, S. J. Hinshaw, S. A. Tycho, D. Pellacani, P. Xiang, K. Muthuraman, S. Sobhanifar, M. H. Piper, F. J. Triana, J. Hendle, A. Pustilnik, A. C. Adams, S. J. Berens, R. S. Baric, D. R. Martinez, R. W. Cross, T. W. Geisbert, V. Borisevich, O. Abiona, H. M. Belli, M. de Vries, A. Mohamed, M. Dittmann, M. I. Samanovic, M. J. Mulligan, J. A. Goldsmith, C.-L. Hsieh, N. V. Johnson, D. Wrapp, J. S. McLellan, B. C. Barnhart, B. S. Graham, J. R. Mascola, C. L. Hansen, and E. Falconer, "The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in nonhuman primates," Sci. Transl. Med., vol. 13, May 2021.

[59] J. Hansen, A. Baum, K. E. Pascal, V. Russo, S. Giordano, E. Wloga, B. O. Fulton, Y. Yan, K. Koon, K. Patel, K. M. Chung, A. Hermann, E. Ullman, J. Cruz, A. Rafique, T. Huang, J. Fairhurst, C. Libertiny, M. Malbec, W.-Y. Lee, R. Welsh, G. Farr, S. Pennington, D. Deshpande, J. Cheng, A. Watty, P. Bouffard, R. Babb, N. Levenkova, C. Chen, B. Zhang, A. Romero Hernandez, K. Saotome, Y. Zhou, M. Franklin, S. Sivapalasingam, D. C. Lye, S. Weston, J. Logue, R. Haupt, M. Frieman, G. Chen, W. Olson, A. J. Murphy, N. Stahl, G. D. Yancopoulos, and C. A. Kyratsous, "Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail," Science, vol. 369, pp. 1010–1014, Aug. 2020.

[60] R. T. Hietpas, J. D. Jensen, and D. N. A. Bolon, "Experimental illumination of a fitness landscape," Proc. Natl. Acad. Sci. U. S. A., vol. 108, pp. 7896–7901, May 2011.

[61] C. A. Kowalsky, M. S. Faber, A. Nath, H. E. Dann, V. W. Kelly, L. Liu, P. Shanker, E. K. Wagner, J. A. Maynard, C. Chan, and T. A. Whitehead, "Rapid fine conformational epitope mapping using comprehensive mutagenesis and deep sequencing," J. Biol. Chem., vol. 290, pp. 26457–26470, Oct. 2015.

[62] T. Van Blarcom, A. Rossi, D. Foletti, P. Sundar, S. Pitts, C. Bee, J. Melton Witt, Z. Melton, A. Hasa-Moreno, L. Shaughnessy, D. Telman, L. Zhao, W. L. Cheung, J. Berka, W. Zhai, P. Strop, J. Chaparro-Riggers, D. L. Shelton, J. Pons, and A. Rajpal, "Precise and efficient antibody epitope determination through library design, yeast display and next-generation sequencing," J. Mol. Biol., vol. 427, pp. 1513–1534, Mar. 2015.

[63] K. M. Doolan and D. W. Colby, "Conformation-dependent epitopes recognized by prion protein antibodies probed using mutational scanning and deep sequencing," J. Mol. Biol., vol. 427, pp. 328–340, Jan. 2015.

[64] K. K. Chan, D. Dorosky, P. Sharma, S. A. Abbasi, J. M. Dye, D. M. Kranz, A. S. Herbert, and E. Procko, "Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2," Science, vol. 369, pp. 1261–1265, Sept. 2020.

[65] T. N. Starr, A. J. Greaney, S. K. Hilton, D. Ellis, K. H. D. Crawford, A. S. Dingens, M. J. Navarro, J. E. Bowen, M. A. Tortorici, A. C. Walls, N. P. King, D. Veesler, and J. D. Bloom, "Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding," Cell, vol. 182, pp. 1295–1310.e20, Sept. 2020.

[66] T. N. Starr, A. J. Greaney, A. S. Dingens, and J. D. Bloom, "Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016," Cell Rep Med, vol. 2, p. 100255, Apr. 2021.

[67] K. K. Chan, T. J. C. Tan, K. K. Narayanan, and E. Procko, "An engineered decoy receptor for SARS-CoV-2 broadly binds protein S sequence variants," Sci Adv, vol. 7, Feb. 2021.

[68] R. M. Adams, T. Mora, A. M. Walczak, and J. B. Kinney, "Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves," Elife, vol. 5, Dec. 2016.

[69] "SARS-CoV-2 variant classifications and definitions." https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html.

[70] "Tracking omicron and other coronavirus variants." https://www.nytimes.com/interactive/2021/health/coronavirus-variant-tracker.html.

[71] G. Chao, W. L. Lau, B. J. Hackel, S. L. Sazinsky, S. M. Lippow, and K. D. Wittrup, "Isolating and engineering human antibodies using yeast surface display," Nat. Protoc., vol. 1, no. 2, pp. 755–768, 2006.

[72] Y. Jigami, "Yeast glycobiology and its application," Biosci. Biotechnol. Biochem., vol. 72, pp. 637–648, Mar. 2008.

[73] E. Firnberg and M. Ostermeier, "PFunkel: efficient, expansive, user-defined mutagenesis," PLoS One, vol. 7, p. e52031, Dec. 2012.

[74] P. J. Steiner, Z. T. Baumer, and T. A. Whitehead, "A method for user-defined mutagenesis by integrating oligo pool synthesis technology with nicking mutagenesis," Bio Protoc, vol. 10, p. e3697, Aug. 2020.

[75] L. R. Baden, H. M. El Sahly, B. Essink, K. Kotloff, S. Frey, R. Novak, D. Diemert, S. A. Spector, N. Rouphael, C. B. Creech, J. McGettigan, S. Khetan, N. Segall, J. Solis, A. Brosz, C. Fierro, H. Schwartz, K. Neuzil, L. Corey, P. Gilbert, H. Janes, D. Follmann, M. Marovich, J. Mascola, L. Polakowski, J. Ledgerwood, B. S. Graham, H. Bennett, R. Pajon, C. Knightly, B. Leav, W. Deng, H. Zhou, S. Han, M. Ivarsson, J. Miller, T. Zaks, and COVE Study Group, "Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine," N. Engl. J. Med., vol. 384, pp. 403–416, Feb. 2021.

[76] F. P. Polack, S. J. Thomas, N. Kitchin, J. Absalon, A. Gurtman, S. Lockhart, J. L. Perez, G. Pérez Marc, E. D. Moreira, C. Zerbini, R. Bailey, K. A. Swanson, S. Roychoudhury, K. Koury, P. Li, W. V. Kalina, D. Cooper, R. W. Frenck, Jr, L. L. Hammitt, Ö. Türeci, H. Nell, A. Schaefer, S. Ünal, D. B. Tresnan, S. Mather, P. R. Dormitzer, U. Şahin, K. U. Jansen, W. C. Gruber, and C4591001 Clinical Trial Group, "Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine," N. Engl. J. Med., vol. 383, pp. 2603–2615, Dec. 2020.

[77] M. Voysey, S. A. C. Clemens, S. A. Madhi, L. Y. Weckx, P. M. Folegatti, P. K. Aley, B. Angus, V. L. Baillie, S. L. Barnabas, Q. E. Bhorat, S. Bibi, C. Briner, P. Cicconi, A. M. Collins, R. Colin-Jones, C. L. Cutland, T. C. Darton, K. Dheda, C. J. A. Duncan, K. R. W. Emary, K. J. Ewer, L. Fairlie, S. N. Faust, S. Feng, D. M. Ferreira, A. Finn, A. L. Goodman, C. M. Green, C. A. Green, P. T. Heath, C. Hill, H. Hill, I. Hirsch, S. H. C. Hodgson, A. Izu, S. Jackson, D. Jenkin, C. C. D. Joe, S. Kerridge, A. Koen, G. Kwatra, R. Lazarus, A. M. Lawrie, A. Lelliott, V. Libri, P. J. Lillie, R. Mallory, A. V. A. Mendes, E. P. Milan, A. M. Minassian, A. McGregor, H. Morrison, Y. F. Mujadidi, A. Nana, P. J. O'Reilly, S. D. Padayachee, A. Pittella, E. Plested, K. M. Pollock, M. N. Ramasamy, S. Rhead, A. V. Schwarzbold, N. Singh, A. Smith, R. Song, M. D. Snape, E. Sprinz, R. K. Sutherland, R. Tarrant, E. C. Thomson, M. E. Török, M. Toshner, D. P. J. Turner, J. Vekemans, T. L. Villafana, M. E. E. Watson, C. J. Williams, A. D. Douglas, A. V. S. Hill, T. Lambe, S. C. Gilbert, A. J. Pollard, M. Aban, F. Abayomi, K. Abeyskera, J. Aboagye, M. Adam, K. Adams, J. Adamson, Y. A. Adelaja, G. Adewetan, S. Adlou, K. Ahmed, Y. Akhalwaya, S. Akhalwaya, A. Alcock, A. Ali, E. R. Allen, L. Allen, T. C. D. S. C. Almeida, M. P. S. Alves, F. Amorim, F. Andritsou, R. Anslow, M. Appleby, E. H. Arbe-Barnes, M. P. Ariaans, B. Arns, L. Arruda, P. Azi, L. Azi, G. Babbage, C. Bailey, K. F. Baker, M. Baker, N. Baker, P. Baker, L. Baldwin, I. Baleanu, D. Bandeira, A. Bara, M. A. S. Barbosa, D. Barker, G. D. Barlow, E. Barnes, A. S. Barr, J. R. Barrett, J. Barrett, L. Bates, A. Batten, K. Beadon, E. Beales, R. Beckley, S. Belij-Rammerstorfer, J. Bell, D. Bellamy, N. Bellei, S. Belton, A. Berg, L. Bermejo, E. Berrie, L. Berry, D. Berzenyi, A. Beveridge, K. R. Bewley, H. Bexhell, S. Bhikha, A. E. Bhorat, Z. E. Bhorat, E. Bijker, G. Birch, S. Birch, A. Bird, O. Bird, K. Bisnauthsing, M. Bittaye, K. Blackstone, L. Blackwell, H. Bletchly, C. L. Blundell, S. R. Blundell, P. Bodalia, B. C. Boettger, E. Bolam, E. Boland, D. Bormans, N. Borthwick, F. Bowring, A. Boyd, P. Bradley, T. Brenner, P. Brown, C. Brown, C. Brown-O'Sullivan, S. Bruce, E. Brunt, R. Buchan, W. Budd, Y. A. Bulbulia, M. Bull, J. Burbage, H. Burhan, A. Burn, K. R. Buttigieg, N. Byard, I. Cabera Puig, G. Calderon, A. Calvert, S. Camara, M. Cao, F. Cappuccini, J. R. Cardoso, M. Carr, M. W. Carroll, A. Carson-Stevens, Y. d. M. Carvalho, J. A. M. Carvalho, H. R. Casey, P. Cashen, T. Castro, L. C. Castro, K. Cathie, A. Cavey, J. Cerbino-Neto, J. Chadwick, D. Chapman, S. Charlton, I. Chelysheva, O. Chester, S. Chita, J.-S. Cho, L. Cifuentes, E. Clark, M. Clark, A. Clarke, E. A. Clutterbuck, S. L. K. Collins, C. P. Conlon, S. Connarty, N. Coombes, C. Cooper, R. Cooper, L. Cornelissen, T. Corrah, C. Cosgrove, T. Cox, W. E. M. Crocker, S. Crosbie, L. Cullen, D. Cullen, D. R. M. F. Cunha, C. Cunningham, F. C. Cuthbertson, S. N. F. Da Guarda, L. P. da Silva, B. E. Damratoski, Z. Danos, M. T. D. C. Dantas, P. Darroch, M. S. Datoo, C. Datta, M. Davids, S. L. Davies, H. Davies, E. Davis, J. Davis, J. Davis, M. De Nobrega, L. M. De Oliveira Kalid, D. Dearlove, T. Demissie, A. Desai, S. Di Marco, C. Di Maso, M. I. S. Dinelli, T. Dinesh, C. Docksey, C. Dold, T. Dong, F. R. Donnellan, T. Dos Santos, T. G. dos Santos, E. P. Dos Santos, N. Douglas, C. Downing, J. Drake, R. Drake-Brockman, K. Driver, R. Drury, S. J. Dunachie, B. S. Durham, L. Dutra, N. J. W. Easom, S. van Eck, M. Edwards, N. J.

Edwards, O. M. El Muhanna, S. C. Elias, M. Elmore, M. English, A. Esmail, Y. M. Essack, E. Farmer, M. Farooq, M. Farrar, L. Farrugia, B. Faulkner, S. Fedosyuk, S. Felle, S. Feng, C. Ferreira Da Silva, S. Field, R. Fisher, A. Flaxman, J. Fletcher, H. Fofie, H. Fok, K. J. Ford, J. Fowler, P. H. A. Fraiman, E. Francis, M. M. Franco, J. Frater, M. S. M. Freire, S. H. Fry, S. Fudge, J. Furze, M. Fuskova, P. Galian-Rubio, E. Galiza, H. Garlant, M. Gavrila, A. Geddes, K. A. Gibbons, C. Gilbride, H. Gill, S. Glynn, K. Godwin, K. Gokani, U. C. Goldoni, M. Goncalves, I. G. S. Gonzalez, J. Goodwin, A. Goondiwala, K. Gordon-Quayle, G. Gorini, J. Grab, L. Gracie, M. Greenland, N. Greenwood, J. Greffrath, M. M. Groenewald, L. Grossi, G. Gupta, M. Hackett, B. Hallis, M. Hamaluba, E. Hamilton, J. Hamlyn, D. Hammersley, A. T. Hanrath, B. Hanumunthadu, S. A. Harris, C. Harris, T. Harris, T. D. Harrison, D. Harrison, T. C. Hart, B. Hartnell, S. Hassan, J. Haughney, S. Hawkins, J. Hay, I. Head, J. Henry, M. Hermosin Herrera, D. B. Hettle, J. Hill, G. Hodges, E. Horne, M. M. Hou, C. Houlihan, E. Howe, N. Howell, J. Humphreys, H. E. Humphries, K. Hurley, C. Huson, A. Hyder-Wright, C. Hyams, S. Ikram, A. Ishwarbhai, M. Ivan, P. Iveson, V. Iyer, F. Jackson, J. De Jager, S. Jaumdally, H. Jeffers, N. Jesudason, B. Jones, K. Jones, E. Jones, C. Jones, M. R. Jorge, A. Jose, A. Joshi, E. A. M. S. Júnior, J. Kadziola, R. Kailath, F. Kana, K. Karampatsas, M. Kasanyinga, J. Keen, E. J. Kelly, D. M. Kelly, D. Kelly, S. Kelly, D. Kerr, R. d. Á. Kfouri, L. Khan, B. Khozoee, S. Kidd, A. Killen, J. Kinch, P. Kinch, L. D. W. King, T. B. King, L. Kingham, P. Klenerman, F. Knapper, J. C. Knight, D. Knott, S. Koleva, M. Lang, G. Lang, C. W. Larkworthy, J. P. J. Larwood, R. Law, E. M. Lazarus, A. Leach, E. A. Lees, N.-M. Lemm, A. Lessa, S. Leung, Y. Li, A. M. Lias, K. Liatsikos, A. Linder, S. Lipworth, S. Liu, X. Liu, A. Lloyd, S. Lloyd, L. Loew, R. Lopez Ramon, L. Lora, V. Lowthorpe, K. Luz, J. C. MacDonald, G. MacGregor, M. Madhavan, D. O. Mainwaring, E. Makambwa, R. Makinson, M. Malahleha, R. Malamatsho, G. Mallett, K. Mansatta, T. Maoko, K. Mapetla, N. G. Marchevsky, S. Marinou, E. Marlow, G. N. Marques, P. Marriott, R. P. Marshall, J. L. Marshall, F. J. Martins, M. Masenya, M. Masilela, S. K. Masters, M. Mathew, H. Matlebjane, K. Matshidiso, O. Mazur, A. Mazzella, H. McCaughan, J. McEwan, J. McGlashan, L. McInroy, Z. McIntyre, D. McLenaghan, N. McRobert, S. McSwiggan, C. Megson, S. Mehdipour, W. Meijs, R. N. Á. Mendonça, A. J. Mentzer, N. Mirtorabi, C. Mitton, S. Mnyakeni, F. Moghaddas, K. Molapo, M. Moloi, M. Moore, M. I. Moraes-Pinto, M. Moran, E. Morey, R. Morgans, S. Morris, S. Morris, H. C. Morris, F. Morselli, G. Morshead, R. Morter, L. Mottal, A. Moultrie, N. Moya, M. Mpelembue, S. Msomi, Y. Mugodi, E. Mukhopadhyay, J. Muller, A. Munro, C. Munro, S. Murphy, P. Mweu, C. H. Myasaki, G. Naik, K. Naker, E. Nastouli, A. Nazir, B. Ndlovu, F. Neffa, C. Njenga, H. Noal, A. Noé, G. Novaes, F. L. Nugent, G. Nunes, K. O'Brien, D. O'Connor, M. Odam, S. Oelofse, B. Oguti, V. Olchawski, N. J. Oldfield, M. G. Oliveira, C. Oliveira, A. Oosthuizen, P. O'Reilly, P. Osborne, D. R. J. Owen, L. Owen, D. Owens, N. Owino, M. Pacurar, B. V. B. Paiva, E. M. F. Palhares, S. Palmer, S. Parkinson, H. M. R. T. Parracho, K. Parsons, D. Patel, B. Patel, F. Patel, K. Patel, M. Patrick-Smith, R. O. Payne, Y. Peng, E. J. Penn, A. Pennington, M. P. Peralta Alvarez, J. Perring, N. Perry, R. Perumal, S. Petkar, T. Philip, D. J. Phillips, J. Phillips, M. K. Phohu, L. Pickup, S. Pieterse, J. Piper, D. Pipini, M. Plank, J. Du Plessis, S. Pollard, J. Pooley, A. Pooran, I. Poulton, C. Powers, F. B. Presa, D. A. Price, V. Price, M. Primeira, P. C. Proud, S. Provstgaard-Morys, S. Pueschel, D. Pulido, S. Quaid, R. Rabara, A. Radford, K. Radia, D. Rajapaska, T. Rajeswaran, A. S. F. Ramos, F. Ramos Lopez, T. Rampling, J. Rand, H. Ratcliffe, T. Rawlinson, D. Rea, B. Rees, J. Reiné, M. Resuello-Dauti, E. Reyes Pabon, C. M. Ribiero, M. Ricamara, A. Richter, N. Ritchie, A. J. Ritchie, A. J. Robbins, H. Roberts, R. E. Robinson, H. Robinson, T. T. Rocchetti, B. P. Rocha,

S. Roche, C. Rollier, L. Rose, A. L. Ross Russell, L. Rossouw, Royal, Simon, I. Rudiansyah, S. Ruiz, S. Saich, C. Sala, J. Sale, A. M. Salman, N. Salvador, S. Salvador, M. Sampaio, A. D. Samson, A. Sanchez-Gonzalez, H. Sanders, K. Sanders, E. Santos, M. F. S. Santos Guerra, I. Satti, J. E. Saunders, C. Saunders, A. Sayed, I. Schim van der Loeff, A. B. Schmid, E. Schofield, G. Screaton, S. Seddiqi, R. R. Segireddy, R. Senger, S. Serrano, R. Shah, I. Shaik, H. E. Sharpe, K. Sharrocks, R. Shaw, A. Shea, A. Shepherd, J. G. Shepherd, F. Shiham, E. Sidhom, S. E. Silk, A. C. da Silva Moraes, G. Silva-Junior, L. Silva-Reyes, A. D. Silveira, M. B. V. Silveira, J. Sinha, D. T. Skelly, D. C. Smith, N. Smith, H. E. Smith, D. J. Smith, C. C. Smith, A. Soares, T. Soares, C. Solórzano, G. L. Sorio, K. Sorley, T. Sosa-Rodriguez, C. M. C. D. L. Souza, B. S. D. F. Souza, A. R. Souza, A. J. Spencer, F. Spina, L. Spoors, L. Stafford, I. Stamford, I. Starinskij, R. Stein, J. Steven, L. Stockdale, L. V. Stockwell, L. H. Strickland, A. C. Stuart, A. Sturdy, N. Sutton, A. Szigeti, A. Tahiri-Alaoui, R. Tanner, C. Taoushanis, A. W. Tarr, K. Taylor, U. Taylor, I. J. Taylor, J. Taylor, R. te Water Naude, Y. Themistocleous, A. Themistocleous, M. Thomas, K. Thomas, T. M. Thomas, A. Thombrayil, F. Thompson, A. Thompson, K. Thompson, A. Thompson, J. Thomson, V. Thornton-Jones, P. J. Tighe, L. A. Tinoco, G. Tiongson, B. Tladinyane, M. Tomasicchio, A. Tomic, S. Tonks, J. Towner, N. Tran, J. Tree, G. Trillana, C. Trinham, R. Trivett, A. Truby, B. L. Tsheko, A. Turabi, R. Turner, C. Turner, M. Ulaszewska, B. R. Underwood, R. Varughese, D. Verbart, M. Verheul, I. Vichos, T. Vieira, C. S. Waddington, L. Walker, E. Wallis, M. Wand, D. Warbick, T. Wardell, G. Warimwe, S. C. Warren, B. Watkins, E. Watson, S. Webb, A. Webb-Bridges, A. Webster, J. Welch, J. Wells, A. West, C. White, R. White, P. Williams, R. L. Williams, R. Winslow, M. Woodyer, A. T. Worth, D. Wright, M. Wroblewska, A. Yao, R. Zimmer, D. Zizi, and P. Zuidewind, "Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in brazil, south africa, and the UK," <u>Lancet</u>, vol. 397, pp. 99–111, Jan. 2021.

[78] P. Wang, M. S. Nair, L. Liu, S. Iketani, Y. Luo, Y. Guo, M. Wang, J. Yu, B. Zhang, P. D. Kwong, B. S. Graham, J. R. Mascola, J. Y. Chang, M. T. Yin, M. Sobieszczyk, C. A. Kyratsous, L. Shapiro, Z. Sheng, Y. Huang, and D. D. Ho, "Antibody resistance of SARS-CoV-2 variants b.1.351 and b.1.1.7," <u>Nature</u>, vol. 593, pp. 130–135, May 2021.

[79] W. Zhang, B. D. Davis, S. S. Chen, J. M. Sincuir Martinez, J. T. Plummer, and E. Vail, "Emergence of a novel SARS-CoV-2 variant in southern california," <u>JAMA</u>, vol. 325, pp. 1324–1326, Apr. 2021.

[80] R. E. Chen, X. Zhang, J. B. Case, E. S. Winkler, Y. Liu, L. A. VanBlargan, J. Liu, J. M. Errico, X. Xie, N. Suryadevara, P. Gilchuk, S. J. Zost, S. Tahan, L. Droit, J. S. Turner, W. Kim, A. J. Schmitz, M. Thapa, D. Wang, A. C. M. Boon, R. M. Presti, J. A. O'Halloran, A. H. J. Kim, P. Deepak, D. Pinto, D. H. Fremont, J. E. Crowe, Jr, D. Corti, H. W. Virgin, A. H. Ellebedy, P.-Y. Shi, and M. S. Diamond, "Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies," <u>Nat. Med.</u>, vol. 27, pp. 717–726, Apr. 2021.

[81] C. W. Tan, W. N. Chia, X. Qin, P. Liu, M. I.-C. Chen, C. Tiu, Z. Hu, V. C.-W. Chen, B. E. Young, W. R. Sia, Y.-J. Tan, R. Foo, Y. Yi, D. C. Lye, D. E. Anderson, and L.-F. Wang, "A SARS-CoV-2 surrogate virus neutralization test based on antibody-mediated blockage

of ACE2–spike protein–protein interaction," <u>Nat. Biotechnol.</u>, vol. 38, pp. 1073–1078, July 2020.

[82] E. E. Wrenbeck, J. R. Klesmith, J. A. Stapleton, A. Adeniran, K. E. J. Tyo, and T. A. Whitehead, "Plasmid-based one-pot saturation mutagenesis," <u>Nat. Methods</u>, vol. 13, pp. 928–930, Nov. 2016.

[83] N. C. Wu, M. Yuan, H. Liu, C.-C. D. Lee, X. Zhu, S. Bangaru, J. L. Torres, T. G. Caniels, P. J. M. Brouwer, M. J. van Gils, R. W. Sanders, A. B. Ward, and I. A. Wilson, "An alternative binding mode of IGHV3-53 antibodies to the SARS-CoV-2 receptor binding domain," <u>Cell Rep.</u>, vol. 33, p. 108274, Oct. 2020.

[84] Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, C. Zhao, Q. Zhang, H. Liu, L. Nie, H. Qin, M. Wang, Q. Lu, X. Li, Q. Sun, J. Liu, L. Zhang, X. Li, W. Huang, and Y. Wang, "The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity," <u>Cell</u>, vol. 182, pp. 1284–1294.e9, Sept. 2020.

[85] Z. Liu, L. A. VanBlargan, L.-M. Bloyet, P. W. Rothlauf, R. E. Chen, S. Stumpf, H. Zhao, J. M. Errico, E. S. Theel, M. J. Liebeskind, B. Alford, W. J. Buchser, A. H. Ellebedy, D. H. Fremont, M. S. Diamond, and S. P. J. Whelan, "Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization," <u>Cell Host Microbe</u>, vol. 29, pp. 477–488.e4, Mar. 2021.

[86] Y. Weisblum, F. Schmidt, F. Zhang, J. DaSilva, D. Poston, J. C. Lorenzi, F. Muecksch, M. Rutkowska, H.-H. Hoffmann, E. Michailidis, C. Gaebler, M. Agudelo, A. Cho, Z. Wang, A. Gazumyan, M. Cipolla, L. Luchsinger, C. D. Hillyer, M. Caskey, D. F. Robbiani, C. M. Rice, M. C. Nussenzweig, T. Hatziioannou, and P. D. Bieniasz, "Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants," <u>Elife</u>, vol. 9, Oct. 2020.

[87] V. Potapov, J. L. Ong, R. B. Kucera, B. W. Langhorst, K. Bilotti, J. M. Pryor, E. J. Cantor, B. Canton, T. F. Knight, T. C. Evans, Jr, and G. J. S. Lohman, "Comprehensive profiling of four base overhang ligation fidelity by T4 DNA ligase and application to DNA assembly," <u>ACS Synth. Biol.</u>, vol. 7, pp. 2665–2674, Nov. 2018.

[88] C. Engler and S. Marillonnet, "Golden gate cloning," <u>Methods Mol. Biol.</u>, vol. 1116, pp. 119–131, 2014.

[89] J. ter Meulen, E. N. van den Brink, L. L. M. Poon, W. E. Marissen, C. S. W. Leung, F. Cox, C. Y. Cheung, A. Q. Bakker, J. A. Bogaards, E. van Deventer, W. Preiser, H. W. Doerr, V. T. Chow, J. de Kruif, J. S. M. Peiris, and J. Goudsmit, "Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants," <u>PLoS Med.</u>, vol. 3, p. e237, July 2006.

[90] A. V. Medina-Cucurella and T. A. Whitehead, "Characterizing Protein-Protein interactions using deep sequencing coupled to yeast surface display," <u>Methods Mol. Biol.</u>, vol. 1764, pp. 101–121, 2018.

[91] E. T. Boder and K. D. Wittrup, "Yeast surface display for screening combinatorial polypeptide libraries," Nat. Biotechnol., vol. 15, pp. 553–557, June 1997.

[92] A. V. Medina-Cucurella, P. J. Steiner, M. S. Faber, J. Beltrán, A. N. Borelli, M. B. Kirby, S. R. Cutler, and T. A. Whitehead, "User-defined single pot mutagenesis using unamplified oligo pools," Protein Eng. Des. Sel., vol. 32, pp. 41–45, Sept. 2019.

[93] L. Benatuil, J. M. Perez, J. Belk, and C.-M. Hsieh, "An improved yeast transformation method for the generation of very large human antibody libraries," Protein Eng. Des. Sel., vol. 23, pp. 155–159, Apr. 2010.

[94] T. A. Whitehead, A. Chevalier, Y. Song, C. Dreyfus, S. J. Fleishman, C. De Mattos, C. A. Myers, H. Kamisetty, P. Blair, I. A. Wilson, and D. Baker, "Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing," Nat. Biotechnol., vol. 30, pp. 543–548, May 2012.

[95] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX, vol. 1-2, pp. 19–25, Sept. 2015.

[96] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," J. Chem. Phys., vol. 79, pp. 926–935, July 1983.

[97] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the amber ff99SB protein force field," Proteins, vol. 78, pp. 1950–1958, June 2010.

[98] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," J. Chem. Phys., vol. 126, p. 014101, Jan. 2007.

[99] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," J. Chem. Phys., vol. 81, pp. 3684–3690, Oct. 1984.

[100] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," J. Appl. Phys., vol. 52, pp. 7182–7190, Dec. 1981.

[101] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," J. Comput. Chem., vol. 18, pp. 1463–1472, Sept. 1997.

[102] R. J. Shalloo, S. J. D. Dann, J.-N. Gruse, C. I. D. Underwood, A. F. Antoine, C. Arran, M. Backhouse, C. D. Baird, M. D. Balcazar, N. Bourgeois, J. A. Cardarelli, P. Hatfield, J. Kang, K. Krushelnick, S. P. D. Mangles, C. D. Murphy, N. Lu, J. Osterhoff, K. Põder, P. P. Rajeev, C. P. Ridgers, S. Rozario, M. P. Selwood, A. J. Shahani, D. R. Symes, A. G. R. Thomas, C. Thornton, Z. Najmudin, and M. J. V. Streeter, "Automation and control of laser wakefield accelerators using bayesian optimization," Nat. Commun., vol. 11, p. 6355, Dec. 2020.

[103] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in Proceedings of the 9th Python in Science Conference, vol. 57, pp. 10–25080, 2010.

[104] C. B. Creech, S. C. Walker, and R. J. Samuels, "SARS-CoV-2 vaccines," JAMA, vol. 325, pp. 1318–1320, Apr. 2021.

[105] L. Piccoli and Y. Park, "J., tortorici MA, czudnochowski n., walls AC, bel-tramello m., et al. mapping neutralizing and immuno-dominant sites on the SARS-CoV-2 spike receptor ...," Cell, 2022.

[106] R. Copin, A. Baum, E. Wloga, K. E. Pascal, S. Giordano, B. O. Fulton, A. Zhou, N. Negron, K. Lanza, N. Chan, A. Coppola, J. Chiu, M. Ni, Y. Wei, G. S. Atwal, A. R. Hernandez, K. Saotome, Y. Zhou, M. C. Franklin, A. T. Hooper, S. McCarthy, S. Hamon, J. D. Hamilton, H. M. Staples, K. Alfson, R. Carrion, Jr, S. Ali, T. Norton, S. Somersan-Karakaya, S. Sivapalasingam, G. A. Herman, D. M. Weinreich, L. Lipsich, N. Stahl, A. J. Murphy, G. D. Yancopoulos, and C. A. Kyratsous, "The monoclonal antibody combination REGEN-COV protects against SARS-CoV-2 mutational escape in preclinical and human studies," Cell, vol. 184, pp. 3949–3961.e11, July 2021.

[107] D. M. Weinreich, S. Sivapalasingam, T. Norton, S. Ali, H. Gao, R. Bhore, B. J. Musser, Y. Soo, D. Rofail, J. Im, C. Perry, C. Pan, R. Hosain, A. Mahmood, J. D. Davis, K. C. Turner, A. T. Hooper, J. D. Hamilton, A. Baum, C. A. Kyratsous, Y. Kim, A. Cook, W. Kampman, A. Kohli, Y. Sachdeva, X. Graber, B. Kowal, T. DiCioccio, N. Stahl, L. Lipsich, N. Braunstein, G. Herman, G. D. Yancopoulos, and Trial Investigators, "REGN-COV2, a neutralizing antibody cocktail, in outpatients with covid-19," N. Engl. J. Med., vol. 384, pp. 238–251, Jan. 2021.

[108] P. Wang, R. G. Casner, M. S. Nair, M. Wang, J. Yu, G. Cerutti, L. Liu, P. D. Kwong, Y. Huang, L. Shapiro, and D. D. Ho, "Increased resistance of SARS-CoV-2 variant p.1 to antibody neutralization," Cell Host Microbe, vol. 29, pp. 747–751.e4, May 2021.

[109] C. K. Wibmer, F. Ayres, T. Hermanus, M. Madzivhandila, P. Kgagudi, B. Oosthuysen, B. E. Lambson, T. de Oliveira, M. Vermeulen, K. van der Berg, T. Rossouw, M. Boswell, V. Ueckermann, S. Meiring, A. von Gottberg, C. Cohen, L. Morris, J. N. Bhiman, and P. L. Moore, "SARS-CoV-2 501Y.V2 escapes neutralization by south african COVID-19 donor plasma," Nat. Med., vol. 27, pp. 622–625, Apr. 2021.

[110] M. Yuan, D. Huang, C.-C. D. Lee, N. C. Wu, A. M. Jackson, X. Zhu, H. Liu, L. Peng, M. J. van Gils, R. W. Sanders, D. R. Burton, S. M. Reincke, H. Prüss, J. Kreye, D. Nemazee, A. B. Ward, and I. A. Wilson, "Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants," Science, vol. 373, pp. 818–823, Aug. 2021.

[111] J. Fedry, D. L. Hurdiss, C. Wang, W. Li, G. Obal, I. Drulyte, W. Du, S. C. Howes, F. J. M. van Kuppeveld, F. Förster, and B.-J. Bosch, "Structural insights into the cross-neutralization of SARS-CoV and SARS-CoV-2 by the human monoclonal antibody 47D11," Sci Adv, vol. 7, June 2021.

[112] D. Li, R. J. Edwards, K. Manne, D. R. Martinez, A. Schäfer, S. M. Alam, K. Wiehe, X. Lu, R. Parks, L. L. Sutherland, T. H. Oguin, 3rd, C. McDanal, L. G. Perez, K. Mansouri, S. M. C. Gobeil, K. Janowska, V. Stalls, M. Kopp, F. Cai, E. Lee, A. Foulger, G. E. Hernandez, A. Sanzone, K. Tilahun, C. Jiang, L. V. Tse, K. W. Bock, M. Minai, B. M. Nagata, K. Cronin, V. Gee-Lai, M. Deyton, M. Barr, T. Von Holle, A. N. Macintyre, E. Stover, J. Feldman, B. M. Hauser, T. M. Caradonna, T. D. Scobey, W. Rountree, Y. Wang, M. A. Moody, D. W. Cain, C. T. DeMarco, T. N. Denny, C. W. Woods, E. W. Petzold, A. G. Schmidt, I.-T. Teng, T. Zhou, P. D. Kwong, J. R. Mascola, B. S. Graham, I. N. Moore, R. Seder, H. Andersen, M. G. Lewis, D. C. Montefiori, G. D. Sempowski, R. S. Baric, P. Acharya, B. F. Haynes, and K. O. Saunders, "In vitro and in vivo functions of SARS-CoV-2 infection-enhancing and neutralizing antibodies," Cell, vol. 184, pp. 4203–4219.e32, Aug. 2021.

[113] H. Liu, N. C. Wu, M. Yuan, S. Bangaru, J. L. Torres, T. G. Caniels, J. van Schooten, X. Zhu, C.-C. D. Lee, P. J. M. Brouwer, M. J. van Gils, R. W. Sanders, A. B. Ward, and I. A. Wilson, "Cross-neutralization of a SARS-CoV-2 antibody to a functionally conserved site is mediated by avidity (preprint)," 2020.

[114] B. Julg, P.-T. Liu, K. Wagh, W. M. Fischer, P. Abbink, N. B. Mercado, J. B. Whitney, J. P. Nkolola, K. McMahan, L. J. Tartaglia, E. N. Borducchi, S. Khatiwada, M. Kamath, J. A. LeSuer, M. S. Seaman, S. D. Schmidt, J. R. Mascola, D. R. Burton, B. T. Korber, and D. H. Barouch, "Protection against a mixed SHIV challenge by a broadly neutralizing antibody cocktail," Sci. Transl. Med., vol. 9, Sept. 2017.

[115] T. N. Starr, A. J. Greaney, A. Addetia, W. W. Hannon, M. C. Choudhary, A. S. Dingens, J. Z. Li, and J. D. Bloom, "Prospective mapping of viral mutations that escape antibodies used to treat COVID-19," Science, vol. 371, pp. 850–854, Feb. 2021.

[116] L. Xu, A. Pegu, E. Rao, N. Doria-Rose, J. Beninga, K. McKee, D. M. Lord, R. R. Wei, G. Deng, M. Louder, S. D. Schmidt, Z. Mankoff, L. Wu, M. Asokan, C. Beil, C. Lange, W. D. Leuschner, J. Kruip, R. Sendak, Y. D. Kwon, T. Zhou, X. Chen, R. T. Bailer, K. Wang, M. Choe, L. J. Tartaglia, D. H. Barouch, S. O'Dell, J.-P. Todd, D. R. Burton, M. Roederer, M. Connors, R. A. Koup, P. D. Kwong, Z.-Y. Yang, J. R. Mascola, and G. J. Nabel, "Trispecific broadly neutralizing HIV antibodies mediate potent SHIV protection in macaques," Science, vol. 358, pp. 85–90, Oct. 2017.

[117] M. Yuan, H. Liu, N. C. Wu, and I. A. Wilson, "Recognition of the SARS-CoV-2 receptor binding domain by neutralizing antibodies," Biochem. Biophys. Res. Commun., vol. 538, pp. 192–203, Jan. 2021.

[118] F. Zhao, M. Yuan, C. Keating, N. Shaabani, O. Limbo, C. Joyce, J. Woehl, S. Barman, A. Burns, X. Zhu, M. Ricciardi, L. Peng, J. Smith, D. Huang, B. Briney, D. Sok, D. Nemazee, J. R. Teijaro, I. A. Wilson, D. R. Burton, and J. G. Jardine, "Broadening a SARS-CoV-1 neutralizing antibody for potent SARS-CoV-2 neutralization through directed evolution." May 2021.

[119] Q. Zhang, B. Ju, J. Ge, J. F.-W. Chan, L. Cheng, R. Wang, W. Huang, M. Fang, P. Chen, B. Zhou, S. Song, S. Shan, B. Yan, S. Zhang, X. Ge, J. Yu, J. Zhao, H. Wang, L. Liu, Q. Lv, L. Fu, X. Shi, K. Y. Yuen, L. Liu, Y. Wang, Z. Chen, L. Zhang, X. Wang, and Z. Zhang,

"Potent and protective IGHV3-53/3-66 public antibodies and their shared escape mutant on the spike of SARS-CoV-2," Nat. Commun., vol. 12, p. 4210, July 2021.

[120] C. Yi, X. Sun, Y. Lin, C. Gu, L. Ding, X. Lu, Z. Yang, Y. Zhang, L. Ma, W. Gu, A. Qu, X. Zhou, X. Li, J. Xu, Z. Ling, Y. Xie, H. Lu, and B. Sun, "Comprehensive mapping of binding hot spots of SARS-CoV-2 RBD-specific neutralizing antibodies for tracking immune escape variants," Genome Med., vol. 13, p. 164, Oct. 2021.

[121] C. M. Haas, I. M. Francino-Urdaniz, P. J. Steiner, and T. A. Whitehead, "Identification of SARS-CoV-2 S RBD escape mutants using yeast screening and deep mutational scanning," STAR Protoc, vol. 2, p. 100869, Dec. 2021.

[122] J. C. Klein, M. J. Lajoie, J. J. Schwartz, E.-M. Strauch, J. Nelson, D. Baker, and J. Shendure, "Multiplex pairwise assembly of array-derived DNA oligonucleotides," Nucleic Acids Res., vol. 44, p. e43, Mar. 2016.

[123] B. P. Kuiper, R. C. Prins, and S. Billerbeck, "Oligo pools as an affordable source of synthetic DNA for cost-effective library construction in protein- and metabolic pathway engineering," Chembiochem, vol. 23, p. e202100507, Apr. 2022.

[124] D. T. Tresnak and B. J. Hackel, "Deep antimicrobial activity and stability analysis inform lysin Sequence–Function mapping," ACS Synth. Biol., vol. 12, pp. 249–264, Jan. 2023.

[125] K. Hiraga and F. H. Arnold, "General method for sequence-independent site-directed chimeragenesis," J. Mol. Biol., vol. 330, pp. 287–296, July 2003.

[126] W. Szybalski, S. C. Kim, N. Hasan, and A. J. Podhajska, "Class-IIS restriction enzymes–a review," Gene, vol. 100, pp. 13–26, Apr. 1991.

[127] C. Engler, R. Kandzia, and S. Marillonnet, "A one pot, one step, precision cloning method with high throughput capability," PLoS One, vol. 3, p. e3647, Nov. 2008.

[128] C. Engler, R. Gruetzner, R. Kandzia, and S. Marillonnet, "Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes," PLoS One, vol. 4, p. e5553, May 2009.

[129] E. Weber, C. Engler, R. Gruetzner, S. Werner, and S. Marillonnet, "A modular cloning system for standardized assembly of multigene constructs," PLoS One, vol. 6, p. e16765, Feb. 2011.

[130] M. E. Lee, W. C. DeLoache, B. Cervantes, and J. E. Dueber, "A highly characterized yeast toolkit for modular, multipart assembly," ACS Synth. Biol., vol. 4, pp. 975–986, Sept. 2015.

[131] J. E. Bird, J. Marles-Wright, and A. Giachino, "A user's guide to golden gate cloning methods and standards," ACS Synth. Biol., vol. 11, pp. 3551–3563, Nov. 2022.

[132] C. Sellmann, L. Pekar, C. Bauer, E. Ciesielski, S. Krah, S. Becker, L. Toleikis, J. Kügler, A. Frenzel, B. Valldorf, M. Hust, and S. Zielonka, "A One-Step process for the construction of phage display scfv and VHH libraries," Mol. Biotechnol., vol. 62, pp. 228–239, Apr. 2020.

[133] K. Chockalingam, Z. Peng, C. N. Vuong, L. R. Berghman, and Z. Chen, "Golden gate assembly with a bi-directional promoter (GBid): A simple, scalable method for phage display fab library creation," Sci. Rep., vol. 10, p. 2888, Feb. 2020.

[134] H. Gerstmans, D. Grimon, D. Gutiérrez, C. Lood, A. Rodríguez, V. van Noort, J. Lammertyn, R. Lavigne, and Y. Briers, "A VersaTile-driven platform for rapid hit-to-lead development of engineered lysins," Sci Adv, vol. 6, p. eaaz1136, June 2020.

[135] P. Coussement, D. Bauwens, J. Maertens, and M. De Mey, "Direct combinatorial pathway optimization," ACS Synth. Biol., vol. 6, pp. 224–232, Feb. 2017.

[136] C. B. Macdonald, D. Nedrud, P. R. Grimes, D. Trinidad, J. S. Fraser, and W. Coyote-Maestas, "DIMPLE: deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology," Genome Biol., vol. 24, p. 36, Feb. 2023.

[137] P. Püllmann, C. Ulpinnis, S. Marillonnet, R. Gruetzner, S. Neumann, and M. J. Weissenborn, "Golden mutagenesis: An efficient multi-site-saturation mutagenesis approach by golden gate cloning with automated primer design," Sci. Rep., vol. 9, p. 10932, July 2019.

[138] I. K. Strawn, P. J. Steiner, M. S. Newton, Z. T. Baumer, and T. A. Whitehead, "A method for generating user-defined circular single-stranded DNA from plasmid DNA using golden gate intramolecular ligation," Biotechnol. Bioeng., vol. n/a, June 2023.

[139] J. Beltrán, P. J. Steiner, M. Bedewitz, S. Wei, F. C. Peterson, Z. Li, B. E. Hughes, Z. Hartley, N. R. Robertson, A. V. Medina-Cucurella, Z. T. Baumer, A. C. Leonard, S.-Y. Park, B. F. Volkman, D. A. Nusinow, W. Zhong, I. Wheeldon, S. R. Cutler, and T. A. Whitehead, "Rapid biosensor development using plant hormone receptors as reprogrammable scaffolds," Nat. Biotechnol., vol. 40, pp. 1855–1861, June 2022.

[140] R. Kucera and E. Cantor, "Breaking through the limitations of golden gate assembly– the Co-Evolution of test systems, engineered enzymes and understanding ligase fidelity," NEB.

[141] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, 3rd, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases," Nat. Methods, vol. 6, pp. 343–345, May 2009.

[142] J. L. Tenthorey, M. Emerman, and H. S. Malik, "Evolutionary landscapes of Host-Virus arms races," Annu. Rev. Immunol., vol. 40, pp. 271–294, Apr. 2022.

[143] Y. Cao, F. Jian, J. Wang, Y. Yu, W. Song, A. Yisimayi, R. An, X. Chen, N. Zhang, Y. Wang, P. Wang, L. Zhao, H. Sun, L. Yu, S. Yang, X. Niu, T. Xiao, Q. Gu, F. Shao, X. Hao, Y. Xu, R. Jin, Z. Shen, Y. Wang, and X. S. Xie, "Imprinted SARS-CoV-2 humoral immunity induces convergent omicron RBD evolution," Nature, vol. 614, pp. 521–529, Dec. 2022.

[144] J. M. Taft, C. R. Weber, B. Gao, R. A. Ehling, J. Han, L. Frei, S. W. Metcalfe, M. D. Overath, A. Yermanos, W. Kelton, and S. T. Reddy, "Deep mutational learning predicts ACE2 binding and antibody escape to combinatorial mutations in the SARS-CoV-2 receptor-binding domain," Cell, vol. 185, pp. 4008–4022.e14, Oct. 2022.

[145] J. Zahradník, S. Marciano, M. Shemesh, E. Zoler, D. Harari, J. Chiaravalli, B. Meyer, Y. Rudich, C. Li, I. Marton, and Others, "SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution," Nature microbiology, vol. 6, no. 9, pp. 1188–1198, 2021.

[146] B. Berkhout and E. Herrera-Carrillo, "SARS-CoV-2 evolution: On the sudden appearance of the omicron variant," J. Virol., vol. 96, p. e0009022, Apr. 2022.

[147] L. B. Shrestha, C. Foster, W. Rawlinson, N. Tedla, and R. A. Bull, "Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: Implications for immune escape and transmission," Rev. Med. Virol., vol. 32, p. e2381, Sept. 2022.

[148] E. Humphris-Narayanan, E. Akiva, R. Varela, S. Ó Conchúir, and T. Kortemme, "Prediction of mutational tolerance in HIV-1 protease and reverse transcriptase using flexible backbone protein design," PLoS Comput. Biol., vol. 8, p. e1002639, Aug. 2012.

[149] Y. Ji, X. Tong, D. Xu, J. Liao, R. V. Davuluri, G.-Y. Yang, and R. K. Mishra, "11 - a new robust classifier to detect hot-spots and null-spots in protein–protein interface: validation of binding pocket and identification of inhibitors in in vitro and in vivo models," in Big Data Analytics in Chemoinformatics and Bioinformatics (S. C. Basak and M. Vračko, eds.), pp. 247–263, Elsevier, Jan. 2023.

[150] G. Yang, C. M. Miton, and N. Tokuriki, "A mechanistic view of enzyme evolution," Protein Sci., vol. 29, pp. 1724–1747, Aug. 2020.

[151] D. L. Trudeau and D. S. Tawfik, "Protein engineers turned evolutionists—the quest for the optimal starting point," Curr. Opin. Biotechnol., vol. 60, pp. 46–52, Dec. 2019.

[152] P. A. Romero and F. H. Arnold, "Exploring protein fitness landscapes by directed evolution," Nat. Rev. Mol. Cell Biol., vol. 10, pp. 866–876, Dec. 2009.

[153] Y. Liang, S. Yang, L. Zheng, H. Wang, J. Zhou, S. Huang, L. Yang, and Y. Zuo, "Research progress of reduced amino acid alphabets in protein analysis and prediction," Comput. Struct. Biotechnol. J., vol. 20, pp. 3503–3510, July 2022.

[154] A. D. Solis, "Amino acid alphabet reduction preserves fold information contained in contact interactions in proteins," Proteins, vol. 83, pp. 2198–2216, Dec. 2015.

[155] D. S. Karlov, S. Sosnin, M. V. Fedorov, and P. Popov, "graphdelta: MPNN scoring function for the affinity prediction of Protein–Ligand complexes," ACS Omega, vol. 5, pp. 5150–5159, Mar. 2020.

[156] V. Upadhyay, C. Patrick, A. Lucas, and K. M. G. Mallela, "Convergent evolution of multiple mutations improves the viral fitness of SARS-CoV-2 variants by balancing positive and negative selection," Biochemistry, vol. 61, pp. 963–980, June 2022.

[157] G. M. Air, "Influenza neuraminidase," Influenza Other Respi. Viruses, vol. 6, pp. 245–256, July 2012.

[158] D. Stadlbauer, X. Zhu, M. McMahon, J. S. Turner, T. J. Wohlbold, A. J. Schmitz, S. Strohmeier, W. Yu, R. Nachbagauer, P. A. Mudd, I. A. Wilson, A. H. Ellebedy, and F. Krammer, "Broadly protective human antibodies that target the active site of influenza virus neuraminidase," Science, vol. 366, pp. 499–504, Oct. 2019.

[159] D. M. Fowler, C. L. Araya, W. Gerard, and S. Fields, "Enrich: software for analysis of protein function by enrichment and depletion of variants," Bioinformatics, vol. 27, pp. 3430–3431, Dec. 2011.

[160] J. R. Klesmith and B. J. Hackel, "Improved mutant function prediction via PACT: Protein analysis and classifier toolkit," Bioinformatics, vol. 35, pp. 2707–2712, Aug. 2019.

[161] P. Kvam, B. Vidakovic, and S.-J. Kim, Nonparametric Statistics with Applications to Science and Engineering with R. John Wiley & Sons, Oct. 2022.

[162] J. R. Klesmith, J.-P. Bacik, E. E. Wrenbeck, R. Michalczyk, and T. A. Whitehead, "Tradeoffs between enzyme fitness and solubility illuminated by deep mutational scanning," Proc. Natl. Acad. Sci. U. S. A., vol. 114, pp. 2265–2270, Feb. 2017.

Appendix   A

Identification of SARS-CoV-2 S RBD escape mutants using yeast screening and deep mutational scanning

Summary This is a method used to identify escape mutants on the SARS-CoV-2 S RBD using a yeast screen combined with deep mutational scanning. Nearly all (over 90%) potential single S RBD escape mutants can be identified for monoclonal antibodies that directly compete with ACE2 for binding. Six to ten antibodies can be assessed in parallel. This method has been shown to determine escape mutants that are consistent with more laborious SARS-CoV-2 pseudo neutralization assays.
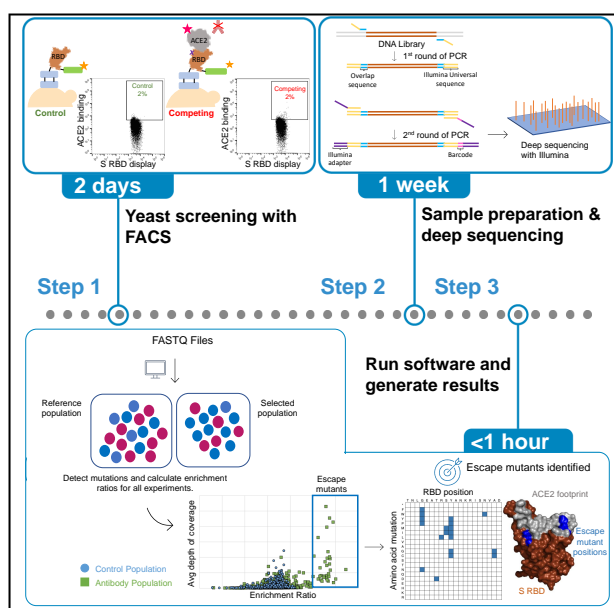
A.1   Before you begin

A.1.1   Protocol overview

This method identifies S RBD nAb escape mutations by identifying S RBD mutants that are not bound by a nAb but retain the ability to bind ACE2. SARS-CoV-2 initiates infection with binding of the Spike receptor binding domain (S RBD) to angiotensin-converting enzyme 2 (ACE2). Many neutralizing antibodies (nAbs) prevent infection by binding to S RBD and blocking ACE2 binding. We have constructed three different S RBD libraries based on the original Wuhan-Hu-1 S RBD N343Q (333-537) (SRA Accession #: SAMN18250431-SAMN18250483). The first is in the original background, the second is in the E484K background, and the third is in the N501Y background. The E484K and N501Y mutations exist in currently circulating Alpha, Beta, and Gamma variants. These libraries are available from AddGene (`https://www.addgene.org/Timothy_Whitehead/`). Each is provided as two sub-libraries (Tile 1 and Tile 2) for compatibility with 250bp paired-end Illumina sequencing. Tile 1 contains mutations at positions 333-437, while tile 2 contains mutations at positions 438-537.

There are three major steps in the protocol and some preparatory work, listed below:

Preparatory Work

A. Prepare chemically competent S. cerevisiae EBY100

B. (i) Plasmid preparation and co-transformation of (ii) yeast backbone and (iii) S RBD libraries into S. cerevisiae EBY100 (**Figure A.1A**)

C. Confirm that neutralizing antibodies selected bind and inhibit the yeast displayed aglycosylated WT RBD (**Figure A.1B-D**)

D. Prepare operating system for running software with its dependencies

Step 1: Yeast screening with FACS

Yeast cells displaying nearly all possible single point mutants on surface exposed positions of S RBD are pre-incubated with a neutralizing antibody before co-incubation with biotinylated ACE2. These cells are then labeled with secondary fluorophores to detect S RBD expression and ACE2 binding. Fluorescence activated cell sorting (FACS) is used to isolate the population of cells still able to bind ACE2.

A. Cell Induction

B. Competitive binding reactions

C. Fluorescent activated cell sorting (FACS)

D. Cell Recovery

Step 2: Sample preparation and deep sequencing

After overnight growth, plasmid DNA from this population is isolated and prepared for deep sequencing.

A. Deep sequencing library preparation

B. Illumina sequencing

Step 3: Run software and generate results

The resulting data from Step 2 are analyzed using custom software that identifies the set of escape mutants.

A.1.2 Preparatory Work A: Prepare chemically competent S. cerevisiae EBY100

Competent yeast can be prepared in advance and stored at -80°C for at least 6 months ([90]).

Timing: 6 hours

1. Prepare chemically competent S. cerevisiae EBY100 (MYA-4941TM)

   (a) Grow EBY100 cells from glycerol stock in 500 mL YPD (recipe) at 30℃ and 300rpm until culture reaches an OD600 of 1.2 (about 4 h).

   (b) Harvest the EBY100 cells by centrifugation at 4000 × g for 5 min.

   (c) Wash the pellet by resuspending in 250 mL sterile H2O and repelleting by centrifugation at 4000 × g for 5 min.

   (d) Wash the pellet with 10 mL of 100 mM lithium acetate and repellet by centrifugation at 4000 × g for 5 min.

   (e) Resuspend cells in 3.5 mL of 100 mM lithium acetate and then add 1.5 mL of 50

   (f) Prepare 210 μL aliquots and store at −80℃. Do not use dry ice to freeze the cells. snap-freeze.

   Note: Typical transformation efficiencies for these prepared cells are 5×10^5 colony forming units (cfu) per μg digested DNA.

A.1.3   Preparatory Work B(i): Plasmid preparation

For this step, the plasmid containing yeast display backbone (pJS697) and a plasmid containing the S RBD insert (see **Table A.1** for variations) are digested. All yeast display plasmids are freely available via AddGene (`https://www.addgene.org/Timothy_Whitehead/`) as E. coli glycerol cell stocks.

Timing: 1 day

Day 1: Plasmid Preparation

Table A.1: List of plasmids that can be used for yeast display screening.

Plasmid	Characteristics	Antibiotic resistance	AddGene collection #
pJS697	Contains yeast display backbone	Kan	`https://www.addgene.org/Timothy_Whitehead/`
pJS699	Contains insert gene encoding Wuhan-Hu-1 S RBD N343Q (333-537) ("WT")	Kan	`https://www.addgene.org/Timothy_Whitehead/`
pIFU001	Contains insert gene encoding Wuhan-Hu-1 S RBD N343Q, E484K (333-537) ("E484K")	Kan	`https://www.addgene.org/Timothy_Whitehead/`
pIFU002	Contains insert gene encoding Wuhan-Hu-1 S RBD N343Q, N501Y (333-537) ("N501Y")	Kan	`https://www.addgene.org/Timothy_Whitehead/`
pJS699_L1	Library of mutations on WT RBD between positions 333 and 437	Kan	`https://www.addgene.org/Timothy_Whitehead/`
pJS699_L2	Library of mutations on WT RBD between positions 438 and 537	Kan	`https://www.addgene.org/Timothy_Whitehead/`
pIFU001_T1	Library of mutations on E484K RBD between positions 333 and 437	Kan	`https://www.addgene.org/Timothy_Whitehead/`
pIFU001_T2	Library of mutations on E484K RBD between positions 438 and 537	Kan	`https://www.addgene.org/Timothy_Whitehead/`
pIFU002_T1	Library of mutations on N501Y RBD between positions 333 and 437	Kan	`https://www.addgene.org/Timothy_Whitehead/`
pIFU002_T2	Library of mutations on N501Y RBD between positions 438 and 537	Kan	`https://www.addgene.org/Timothy_Whitehead/`

2. Digest backbone (pJS697; two BsaI restriction sites) and RBD-encoding (pJS699; two NotI restriction sites) plasmids. The pIFU-series plasmids follow the same procedure as pJS699.

   (a) Set up digestion reactions as follows:

pJS697 (pJS697 for library)	1 μg (5 μg)
10X CutSmart Buffer	5 μL
BsaI-HFv2	1 μL
Nuclease-free water	to 50 μL
Total volume	50 μL

pJS699 (or pJS699_L1, pJS699_L2)	1 μg (5 μg)
10X CutSmart Buffer	5 μL
NotI-HF	1 μL
Nuclease-free water	to 50 μL
Total volume	50 μL

   Note: Digestions are set up following NEB recommendations (NEB cloner)

   (b) Incubate the digestion reactions for one hour at 37℃.

   (c) Purify linear DNA from digested plasmids.

   (i) Purify the backbone from the pJS697 digestion using a Monarch PCR & DNA purification kit. The digestion will result in two linear DNA pieces: the insert between BsaI sites is 26bp and the backbone is 6042bp in length. DNA smaller than 45bp will not stick to the column (Monarch PCR & DNA Cleanup kit) so the eluate will only contain the backbone. Thus, gel electrophoresis and extraction are not necessary with the pJS697 digestion.

   (ii) Purify the linear insert from the pJS699 digestion containing S RBD by gel

electrophoresis and gel extraction. The digestion will result in two linear DNA pieces: one at 832 bp and the other at 2692 bp. Excise the 832 bp fragment and purify using a Monarch Gel Extraction kit (NEB).

(d) Quantify the linear DNA products by A_{260} using a UV-Vis plate reader or nanodrop spectrophotometer. Typical yields for the digestion of 1 µg plasmid DNA are 600 ng (30 ng/µL) for backbone and 100 ng (5 ng/µL) for the gene encoding S RBD.

Note: Linearized plasmids can be stored frozen at -20°C for at least 3 weeks.
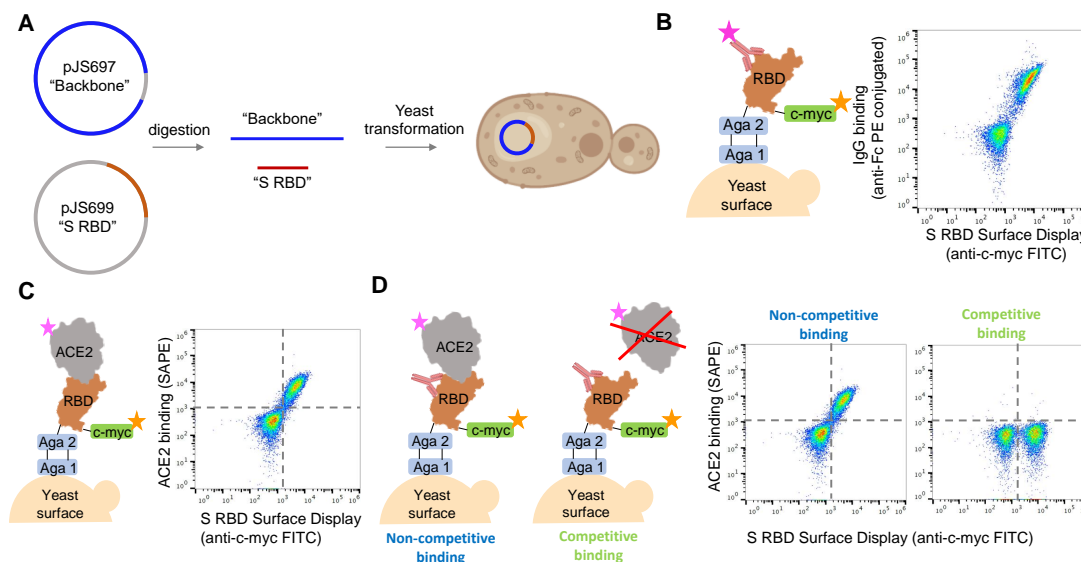
Figure A.1: **Preparative steps for the escape mutant protocol. A:** Homologous recombination in yeast. Left to right: the two plasmids (one containing the backbone and the other one the S RBD insert) are digested independently and the segments are collected. The linearized plasmids are co-transformed into yeast. Yeast cells generate a single plasmid through homologous recombination. **B:** S RBD displayed on the yeast surface bound to antibody. A PE fluorescence signal increase is observed when an antibody binds to S RBD. Two independent populations are detected when screening using FACS. The IgG^{low}/RBD Display^{low} population are yeast cells not containing plasmid. **C:** S RBD labeled with biotinylated ACE2. **D:** Antibodies may bind competitively or non-competitively. Incubation of unlabeled antibody followed by co-incubation with biotinylated ACE2 results in loss of ACE2 binding signal only for the competitive inhibition case. This protocol is designed to identify escape mutants for neutralizing antibodies that function by competitive inhibition of ACE2.

A.1.4   Preparatory Work 1B(ii): Co-transformation of S RBD wild type into S. cerevisiae EBY100

Here, the yeast display backbone from the digest reaction (pJS697) is recombined with the insert containing the S RBD by co-transformation into chemically competent S. cerevisiae EBY100 (ATCC MYA-4941™) following Medina-Cucurella and Whitehead (2018). This step, along with transformations for the S RBD libraries (Preparatory Work 1B(iii)) can be prepared in parallel.

Timing: 3-4 days

Day 1: Plasmid Co-transformation for Linear DNA of Yeast Display Backbone

3. Co-transform linear DNA corresponding to the yeast display backbone and S RBD insert into chemically competent S. cerevisiae EBY100 (MYA-4941™) (**Figure A.1A**).

   (a) Boil 10 μL salmon sperm DNA 10min at 97°C using a Thermal Cycler.

   (b) Thaw chemically competent EBY100. Thawing can be performed either at room temperature (20-25°C) or on ice.

   (c) In a 1.5mL microcentrifuge tube add in order:

EBY100 cells	70 μL
50% w/v PEG 3350	240 μL
1M lithium acetate	36 μL
Salmon sperm DNA	10 μL
Total volume	**356 μL**

   (d) Swirl with pipette tip and vortex to mix thoroughly until solution looks homogeneous. Yeast cells are hardy and will remain viable after a hard mix.

   (e) To 50 μL of the cell mixture, add DNA in a 3:1 molar ratio of insert to backbone for

A.1.5 Preparatory Work 1B(iii): Co-transformation of S RBD libraries into S. cerevisiae EBY100

Timing: 4 days

Day 1: Plasmid Co-transformation for S RBD Libraries

4. Co-transform the libraries into chemically competent S. cerevisiae EBY100 (MYA-4941™) (**Figure A.1A**).

   (a) Boil 30 µl salmon sperm DNA 10min at 97°C using a Thermal Cycler.

   (b) Thaw chemically competent EBY100 at room temperature (20-25°C) or on ice.

   (c) In a 1.5mL microcentrifuge tube add in order:

EBY100 cells	210 μL
50% w/v PEG 3350	720 μL
1M lithium acetate	108 μL
Salmon sperm DNA	30 μL
Total volume	**1068 μL**

   (d) Swirl with pipette tip and vortex to mix thoroughly until solution looks homogeneous. Yeast cells are hardy and will remain viable.

   (e) Add DNA in a 3:1 insert to backbone molar ratio to give a final concentration of 1.5 µg DNA (1.06 µg vector and 440 ng insert). Alternatively, to obtain more transformants, the amount of DNA transformed can increase up to 5 μg total without increasing the amount of other reagents.

   (f) Incubate 30min at 30°C in a static incubator (preferred) or Thermal Cycler.

   (g) Incubate 20min at 42°C in a water bath (preferred) or Thermal Cycler.

(h) Spin down 30s at 16000 $\times$ g using a microcentrifuge. Remove supernatant.

(i) Resuspend in 1mL SDCAA (recipe - make SDCAA$^+$ without antibiotics) and incubate for 5 min at room temperature (20-25°C).

(j) Make 6 serial dilutions and plate 10 μL of each in an SDCAA agar plate to calculate transformation efficiency. Incubate 2-3 days at 30°C. Obtaining near-complete (>99.9%) coverage of the libraries requires $> 1.2 \times 10^5$ transformants for each library ([74]). The first tile (333-437) contains 1120 mutations and the second tile encodes 1260 mutations. The number of transformants obtained can be increased by scaling up the transformation, using more DNA, or using freshly prepared competent cells before freezing.

(k) Transfer remaining culture into 100mL SDCAA$^+$ (recipe) in a 500 ml flask. Incubate at 30°C and 300rpm in the incubator shaker until saturation. This typically takes 48 hours.

Day 4: Making Yeast Stocks

(l.) Make 1 mL yeast stocks at OD$_{600}$=10

    (i) Centrifuge the cell culture 3 min at 3200 $\times$ g.

    (ii) Remove the supernatant and resuspend the cell pellet in yeast storage buffer (recipe) at OD$_{600}$=10. Stocks can be stored for over a year at -80°C.

A.1.6   Preparatory Work C: Confirm that neutralizing antibodies selected bind and inhibit the yeast displayed aglycosylated WT RBD

A recommended though not essential step is to confirm that your neutralizing antibody binds to displayed RBD and can competitively inhibit binding of surface-displayed S RBD to soluble

ACE2. Here, yeast cells are grown from frozen stocks and induced overnight. For labeling, our lab prepares chemically biotinylated soluble ACE2-Fc produced by the Institute of Protein Design ([32]). However, biotinylated soluble ACE2 is now commercially available (e.g. R&D Systems Cat# BT933-020). We have not validated the use of other sources of ACE2 in this protocol. We biotinylate ACE2-Fc with NHS-Biotin following the vendor's protocol (User guide: EZ-link NHS Biotin) using a biotin to protein molar ratio from 20:1 to 80:1.

Timing: 2 days

Day 1: Cell Induction

5. Thaw a 1mL stock of yeast cells harboring the S RBD display plasmid.

6. Pellet the cells by centrifuging 1 min at $16000 \times g$ using a microcentrifuge, remove supernatant, and resuspend in 1 mL SDCAA$^+$ (recipe). Transfer culture to a 14 mL culture tube (VWR Cat#10127-334).

7. Incubate at least 4h at 30°C and 300rpm in an incubator shaker. The cell concentration should double, but the culture should not be saturated. It is convenient to set this culture up in the morning so that cytometer screening can be performed during daytime hours on the following day.

8. Measure the culture OD$_{600}$ with a spectrophotometer. Centrifuge the cells 1 min at $16000 \times g$ in a microcentrifuge and resuspend to OD$_{600}$=1 in SGCAA$^+$ (recipe).

9. Transfer 1 mL of the culture to a new 14 mL culture tube and incubate for 22 h at 22°C and 300 rpm in an incubator shaker.

Day 2: Test cells for (i.) surface display of S RBD; (ii.) ability to bind biotinylated ACE2; (iii.) ability to bind antibody of interest; and (iv.) for antibody ability to competitively inhibit ACE2.

10. Transfer the culture to a 1.5 mL microcentrifuge tube and centrifuge for 1 min at $16000 \times g$ using a microcentrifuge.

11. Measure the culture OD_{600} and resuspend the cells in ice cold PBSF to an $OD_{600}=2$. From this point, keep the cells on ice at all times except when otherwise specified. We typically perform a 10-fold dilution of the culture with PBSF to determine the OD_{600}. The yeast culture does not usually grow during the overnight induction with common resuspension volumes under 1 mL.

   CRITICAL: On the same day, prepare PBSF (recipe recipe) and keep on ice for the duration of the experiments. PBSF needs to be ice cold whenever added to cells.

Screen Unlabeled Cells

12. In a 1.5 mL microcentrifuge tube, add 5 µL cells with 45 µL of PBSF ('unlabeled' cells). Incubate 30 min at RT (20-25°C).

13. Centrifuge 1 min at $16000 \times g$ using a microcentrifuge and resuspend the cells in $100 \mu L$ ice cold PBSF. Screen 25,000 unlabeled cells. These unlabeled cells are used to set a FSC-A/FITC$^+$ gate to collect 0% of the events. Refer to FACS for more detail on setting the gates (**Figure A.2B**).

Check Display of S RBD

14. Label 5 µL cells with 44 μL of PBSF and 1 μL of anti-c-myc FITC in a 1.5mL microcentrifuge tube. Mix by pipetting up and down several times. Incubate the cells on ice for 10 min protected from light.

15. Centrifuge the cells 1 min at $16000 \times g$ using a microcentrifuge. Wash twice with 200 μL ice cold PBSF and remove supernatant. Cell pellets can be stored on ice until ready to screen.
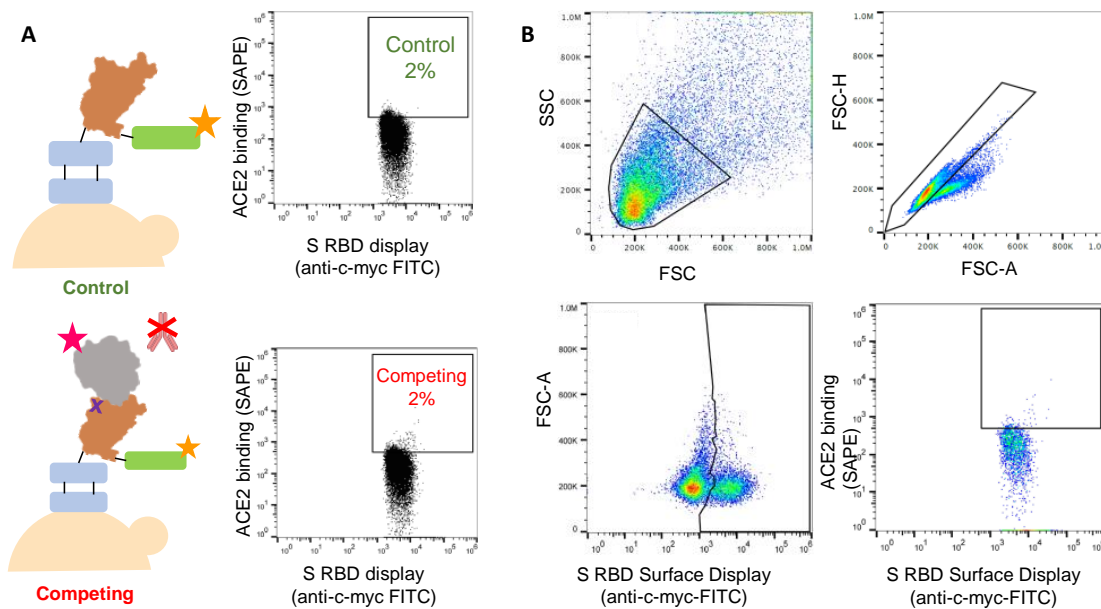
Figure A.2: **Sort gates needed for FACS screening of potential S RBD escape mutants.
A:** Competitive binding reaction cartoons between neutralizing antibody and ACE2. Top is the control where the displayed S RBD is not labelled, top right shows the SAPE$^+$/FITC$^+$ gate collecting the top 2%. Bottom shows escape mutants that no longer bind to nAb but keep the affinity for ACE2. Bottom right shows the SAPE$^+$/FITC$^+$ gate collecting the top 2%. **B:** Collection gates on FACS. FSC/SSC$^+$ gate for isolation of yeast cells, FSC-H/FSC-A gate to discriminate single cells, an FSC-A/FITC$^+$ gate selects the cells displaying the RBD on their surface and from this last gate, the top 2% by a PE$^+$/FITC$^+$ is collected.

16. Resuspend the cells in 100μL ice cold PBSF and screen 25,000 cells using FACS and analyze the FSC-H/FITC$^+$ gate.

    CRITICAL: If less than 30% of cells display RBD, do not continue and induce a new cell stock. Refer to Troubleshooting 1 for potential fixes.

Check ACE2 Binding, Ab Binding, and Competitive Inhibition

17. Set up three separate labeling reactions in three different 1.5 mL microcentrifuge tubes as shown below. For tubes 1 and 3, a lower ACE2 concentration can be used if saturating conditions are maintained (we have also used 30 nM biotinylated ACE2).

	Tube 1	Tube 2	Tube 3
Cells	5μL	5μL	5μL
Biotinylated ACE2	75nM	-	75nM
Antibody	10μg/mL	10μg/mL	10μg/mL
PBSF	To 50μL total volume	To 50μL total volume	To 50μL total volume
	Incubate for 30 minutes at room temperature (20-25°C).		
Biotinylated ACE2	-	-	75nM
	-	-	Incubate for 30 minutes at room temperature (20-25°C).

Optional: Yeast surface titration can be performed following Chao et al. (2006) to determine the effective dissociation constant. The observed dissociation constants of different ACE2-Fc preparations range from 200 pM to 2 nM.

18. Centrifuge 1 min at 16000 x g using a microcentrifuge. Wash twice with 200 µL PBSF and remove supernatant.

19. Resuspend each cell pellet in 50 µL fluorescent labelling mix. The set up for each tube is shown in the following table.

	Tube 1	Tube 2	Tube 3
anti-c-myc FITC	0.6μL	0.6μL	0.6μL
Goat anti-Human IgG FC PE	-	0.25μL	-
Streptavidin-Phycoerythrin (SAPE)	0.25μL	-	0.25μL
PBSF	49.15 μL	49.15 μL	49.15 μL

CRITICAL: Keep the fluorophores on ice and protected from light at all times.

Alternatives: Goat anti-Human IgG Fc PE conjugate is used as the secondary label for the antibody labeling reaction because our neutralizing antibodies are human IgGs. Different antibody platforms (e.g. Fabs, scFvs) or non-human antibodies would require a different secondary label.

20. Incubate 10 min on ice protected from light.

21. Centrifuge 1 min at $16000 \times g$ using a microcentrifuge. Wash twice with 200 μL PBSF and remove supernatant.

    Note: The cell pellet can be stored on ice until ready to screen. We do not recommend storing on ice for longer than 1h as fluorophores can dissociate, resulting in a lower signal.

22. Resuspend the cells in 100 μL PBSF. Screen 25,000 cells.

    CRITICAL: The results of this experiment tell us the likelihood of success for the escape mutant protocol. First, the ACE2 binding population signal must be significantly higher and independent from the non-displaying population on the PE$^+$/FITC$^+$ (**Figure A.1C**). Refer to Troubleshooting 2 to help improve signal. Second, if no PE signal increase is observed for the biotinylated ACE2 labeling (tube 1) **and** antibody labeling (tube 2) yeast cells, then the antibody does not bind to the aglycosylated RBD and cannot be used with this method (see **Figure A.1B** for expected results).

Third, the competitive inhibition reaction should result in minimal PE signal corresponding to ACE2 binding (tube 3). If a PE signal increase is observed, the antibody does not compete with ACE2 for binding to the RBD and thus, cannot be used with this method (**Figure A.1D**).

Alternatives: The fluorescent channel associated with the binding event will be dependent on the choice of secondary label. We use phycoerythrin (PE) as a fluorophore as it is compatible with anti-c-myc FITC and utilizes a 488 nm laser. Other fluorophores can be used.

23. If the antibody binds to RBD and blocks ACE2 binding, co-transform the libraries into yeast and make yeast stocks as described in "Preparation of plasmids and co-transformation into yeast". Store stocks at -80°C.

A.1.7 Preparatory Work D: Prepare operating system for running software with its dependencies

Timing: 20 min

To identify potential escape mutations from deep sequencing data, we developed Python software as outlined in Francino-Urdaniz et al. (2021) [3]. The steps below identify prerequisites necessary for running the software.

24. Download and install Python3 if it is not already installed. Instructions for this process can be found at `https://www.python.org/downloads/`. We recommend using Python v3.9 and higher. While older releases may work, the software has only been tested on recent versions of Python.

25. Install necessary packages.

    (a) Ensure pip is installed with Python3 with the following command at the command

line: pip3 –version. Pip is often installed with newer versions of Python. If pip is not downloaded, further instructions for installation can be found at `https://pip.pypa.io/en/stable/installing/`.

(b) Use pip to download the necessary packages: matplotlib, numpy, openpyxl, pandas, scipy, and statsmodels. These packages can be installed easily at the command line with pip3 –install $package_name$. More detailed instructions are at `https://packaging.python.org/tutorials/installing-packages/`. Download our software available on GitHub:

`https://github.com/WhiteheadGroup/SpikeRBDStabilization.git`. A complete description of the software and its structure are provided in step-by-step methods.

A.2   Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-c-myc FITC, 0.6 μL per 1×10^5 cells	Miltenyi Biotec	Cat# 130-116-485
Goat anti-Human IgG Fc Secondary antibody, PE, eBiosciences, 0.25 μL per 1×10^5 cells	Invitrogen	Cat# 12-4998-82
Bacterial and virus strains		
E. coli Mach1™	Thermo Scientific	Cat# C862003
Biological samples		
N/A		
Chemicals, peptides, and recombinant proteins		

Key Resources Table Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ACE2-Fc	Institute of Protein Design, Laboratory of Prof. Neil King	Walls et al., 2020[32]
NotI-HF	NEB	Cat# R3189
BsaI-HFv2	NEB	Cat# R3733
CutSmart® Buffer	NEB	Cat# B7204S
Nuclease Free Water	IDT	Cat# 11-05-01-14
Ultrapure Agarose	Invitrogen	Cat# 16500-500
TAE Buffer (Tris-acetate-EDTA) (50X)	Thermo Scientific	Cat# B49
SYBR Safe DNA gel stain	Invitrogen	Cat# S33102
Gel loading dye, Purple	NEB	Cat# B7024
UltraPure™ Salmon Sperm DNA Solution	Invitrogen	Cat# 15632011
Glycerol	Macron™ Chemicals	Cat# 5092-16
HEPES free acid	Millipore	Cat# 391338
HEPES sodium salt	Amresco	Cat# 0485
PEG 3350	Spectrum	Cat #P0125
Lithium acetate dihydrate	Sigma-Aldrich	Cat# L6883
PBS - Phosphate-Buffered Saline (10X) pH 7.4	Invitrogen	Cat# AM9624
Streptavidin phycoerythrin (SAPE)	Invitrogen	Cat# S866
Bovine Serum Albumina (BSA), Fraction V, Fatty acid free	VWR	Cat# 7907-25
EZ-Link NHS-Biotin	ThermoFisher	Cat# 20217
Zymolyase	Zymo Research	Cat# E1005
Sodium Chloride	Sigma-Aldrich	Cat# 746398

Key Resources Table Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Difico yeast nitrogen base without amino acids	Sigma-Aldrich	Cat#Y026
Bacto casamino acids, technical grade	Fisher	Cat# 223120
Sodium Phosphate dibasic anhydrous	Fisher Chemical	Cat# BP3321
Sodium Phosphate monobasic monohydrate	Fisher Chemical	Cat# S369-500
D-Galactose Fisher	Bioreagents	Cat# BP656-500
Dextrose	Fisher Chemical	Cat# D19212
Peptone	Fisher	Cat# 211677
Yeast extract	Fisher	Cat# 212750
Agar	BD Biosciences	Cat# 214010
Pen/Strep	Fisher	Cat# 15140-122
Kanamycin	GoldBio	Cat# K-120-25
Exonuclease I	NEB	Cat# M0293S
Lambda exonuclease	NEB	Cat# M0262S
Lambda exonuclease reaction buffer 10X	NEB	Cat# B0262S
Q5 HotStart 2X MasterMix	NEB	Cat# M0494L
rSAP	NEB	Cat# M0371L
70% v/v Denatured ethanol solution	Fisher Bioreagents	Cat# BP82031GAL
IDTE pH 8.0 (1X TE Solution)	IDT	Cat# 11-05-01-13
Quant-iT™PicoGreen™dsDNA Reagent	Thermo Scientific	Cat# T7581
Lambda DNA	Thermo Scientific	Cat# SD0011
Critical commercial assays		
Monarch PCR & DNA Cleanup kit	NEB	Cat# T1030
Monarch DNA Gel Extraction kit	NEB	Cat# T1020
Monarch Plasmid Miniprep kit	NEB	Cat# T1010

Key Resources Table Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Zymoprep Yeast Plasmid miniprep II	Zymo Research	Cat# D2004
Agencourt AMPure XP	Beckman Coulter	Cat# A63881
PhiX	Illumina	Cat# FC-110-3001
Deposited data		
N/A		
Experimental models: Cell lines		
N/A		
Experimental models: Organisms/strains		
Saccharomyces cerevisiae EBY100	ATCC	MYA-4941™
Oligonucleotides		
IFU-104	Francino-Urdaniz et al., 2021[3]	gttcagagttctacagtccgacgatctggaggaggctctgg
IFU-105	Francino-Urdaniz et al., 2021[3]	ccttggcacccgagaattccaccaagctataacgcagcc
IFU-106	Francino-Urdaniz et al., 2021[3]	gttcagagttctacagtccgacgatcggctgcgttatagcttgg
IFU-107	Francino-Urdaniz et al., 2021[3]	ccttggcacccgagaattccagcccccctttgtttttaaccaa
Forward Outer Primer	Kowalsky et al., 2015[61]	aatgatacggcgaccaccgagatctacacgttcagagttctacagtccga

Key Resources Table Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Reverse Outer Primer	Kowalsky et al., 2015[61]	caagcagaagacggcata cgagatnnnnnngtgact ggagttccttggcacccga gaattcca
Recombinant DNA		
pJS697	Francino-Urdaniz et al., 2021[3]	`https://www. addgene. org/Timothy_ Whitehead/`
pJS699	Banach et al. 2021[2]	`https://www. addgene. org/Timothy_ Whitehead/`
pIFU001	Francino-Urdaniz et al., 2021[3]	`https://www. addgene. org/Timothy_ Whitehead/`
pIFU002	Francino-Urdaniz et al., 2021[3]	`https://www. addgene. org/Timothy_ Whitehead/`
pJS699_L1	Francino-Urdaniz et al., 2021[3]	`https://www. addgene. org/Timothy_ Whitehead/`

Key Resources Table Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pJS699_L2	Francino-Urdaniz et al., 2021[3]	`https://www.` `addgene.` `org/Timothy_` `Whitehead/`
pIFU001_T1	Francino-Urdaniz et al., 2021[3]	`https://www.` `addgene.` `org/Timothy_` `Whitehead/`
pIFU001_T2	Francino-Urdaniz et al., 2021[3]	`https://www.` `addgene.` `org/Timothy_` `Whitehead/`
pIFU002_T1	Francino-Urdaniz et al., 2021[3]	`https://www.` `addgene.` `org/Timothy_` `Whitehead/`
pIFU002_T2	Francino-Urdaniz et al., 2021[3]	`https://www.` `addgene.` `org/Timothy_` `Whitehead/`
Software and algorithms		
Python software (dms and analysis modules)	Francino-Urdaniz et al., 2021[3]	`https://` `github.com/` `WhiteheadGroup.` `git`

Key Resources Table Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Python3	N/A	`https://www.python.org/`
Benchling	N/A	`https://www.benchling.com`
Other		
-20°C freezer	VWR	N/A
-80°C freezer	Fisher Scientific, model: REVCO EXF	N/A
Pipettes	N/A	N/A
Centrifuge (doesn't need to be refrigerated)	Eppendorf, model: 5810R	Cat# 05-413-113
Microcentrifuge (doesn't need to be refrigerated)	Fisher, model: accuSpin microcentrifuge 17	Cat# 13-100-675
Vortex Mixer	Thermo Scientific, model: M37165	Cat# M16710-33Q
Static Incubator	VWR, model: Gr Con 4CF	Cat# 89511-420
Incubator shaker	Eppendorf, model: New Brunswick I26 Inc Shaker	Cat# M1324-0000
UV-Vis absorbance plate reader	BioTek, model: Synergy H1M	N/A

Key Resources Table Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Spectrophotometer	Thermo Fisher Scientific, model: 4001/4	Cat# 4001
Cell sorter with a 488 nm laser	SONY, model: LE-SH800SAP	N/A
Thermal Cycler	Eppendorf, model: MastercyclerTM pro	Cat# 950040025
Horizontal Gel Electrophoresis System	BioRad, model: Wide Mini-Sub Cell GT with PowerPac Basic Power Supply	Cat# 1640301

Materials and Equipment

Equipment

(1) -20°C freezer (e.g., VWR)

(2) -80°C freezer (e.g, Fisher Scientific, model: REVCO EXF)

(3) Pipettes

(4) Centrifuge (e.g., Eppendorf, model: 5810R, Cat# 05-413-113; does not need to be refrigerated)

(5) Microcentrifuge (e.g., Fisher, model: accuSpin microcentrifuge 17 , catalog number 13-100-675 ; does not need to be refrigerated)

(6) Vortex Mixer (e.g., Thermo Scientific, model: M37165, Cat# M16710-33Q)

(7) Static Incubator (e.g., VWR, model: Gr Con 4CF, Cat# 89511-420)

(8) Incubator shaker (e.g., Eppendorf, model: New Brunswick I26 Inc Shaker, catalog number: M1324-0000)

(9) UV-Vis absorbance plate reader (e.g., BioTek, model: Synergy H1M)

(10) Spectrophotometer (e.g., Thermo Fisher Scientific, model: 4001/4, catalog number: 4001)

(11) Cell sorter with a 488 nm laser (e.g., SONY, model: LE-SH800SAP)

(12) Thermal Cycler (e.g., Eppendorf, model: Mastercycler™pro, catalog number: 950040025)

(13) Horizontal Gel Electrophoresis System (e.g., BioRad, model: Wide Mini-Sub Cell GT with PowerPac Basic Power Supply, catalog number: 1640301)

A.3   Recipes

Yeast extract-peptone-dextrose (YPD) media

Reagent	Final concentration	Amount
Yeast extract	10 g/L	10 g
Peptone	20 g/L	20 g
Dextrose	20 g/L	20 g
ddH_2O	n/a	to 1 L
Total	**n/a**	**1 L**

Filter sterilize using a 0.22 μm filter. Store at room temperature (20-25°C), protected from light for up to one month.

Yeast storage buffer

Reagent	Final concentration	Amount
Glycerol $\geq$ 99.5%	20% v/v	200 mL
HEPES	10 mM	2.38 g
HEPES sodium salt	10 mM	2.60 g
NaCl	200 mM	11.69 g
ddH_2O	n/a	to 1 L
Total	**n/a**	**1 L**

Take to pH7.5 with NaOH. Filter sterilize using a 0.22 μm filter and store at room temperature (20-25°C) for up to one year.

SCAA media

Reagent	Final concentration	Amount
Difico yeast nitrogen base	6.7 g/L	3.35 g
Bacto casamino acids	5 g/L	2.5 g
$NaH_2PO_4 \cdot H_2O$	62 mM	4.28 g
Na_2HPO_4	38 mM	2.7 g
ddH_2O	n/a	to 450 mL
Total	**n/a**	**450 mL**

Filter sterilize using a 0.22 μm filter and store at room temperature (20-25°C) for up to six months.

CRITICAL: Add NaH2PO4·H2O before Na2HPO4 to ensure full dissolution of the phosphates.

20% w/v Dextrose

Reagent	Final concentration	Amount
Dextrose	20% w/v	100 g
ddH_2O	n/a	to 500 mL
Total	**n/a**	**500 mL**

Filter sterilize using a 0.22μm filter and store at room temperature (20-25°C) for up to six months.

20% w/v Galactose

Reagent	Final concentration	Amount
Galactose	20% w/v	100 g
ddH_2O	n/a	to 500 mL
Total	**n/a**	**500 mL**

Filter sterilize using a 0.22μm filter and store at room temperature (20-25°C) for up to six months.

SDCAA$^+$

Reagent	Final concentration	Amount
20% w/v dextrose	20g/L	10 mL
SCAA	n/a	90 mL
100x Pen/Strep	1X	1 mL
50 mg/mL Kanamycin	50 μg/mL	100 μL
Total	**n/a**	**101.1 mL**

Make on the same day of use and keep at 4°C.

SGCAA$^+$

Reagent	Final concentration	Amount
20% w/v galactose	20g/L	10 mL
SCAA	n/a	90 mL
100x Pen/Strep	1X	1 mL
50 mg/mL Kanamycin	50 μg/mL	100 μL
Total	**n/a**	**101.1 mL**

Make on the same day of use and keep at 4°C.

SDCAA Agar plates Solution A

Reagent	Final concentration	Amount
$NaH_2PO_4 \cdot H_2O$	62 mM	4.28 g
Na_2HPO_4	38 mM	2.7 g
Agar	15 g/L	7.5 g
ddH_2O	n/a	to 450 mL
Total	**n/a**	**450 mL**

Autoclave 20 min at 121°C and 0.5 bar.

Solution B

Reagent	Final concentration	Amount
Difico yeast nitrogen base	6.7 g/L	3.35 g
Bacto casamino acids	5 g/L	2.5 g
Dextrose	20% w/v	10 g
ddH_2O	n/a	to 50 mL
Total	**n/a**	**50 mL**

Filter sterilize using a 0.22μm filter.

Cool autoclaved mixture (**Solution A**) with stirring until below 50°C, add filter-sterilized

solution (**Solution B**), mix well, and pour plates. Plates can be stored for up to 6 months at 4°C.

PBSF

Reagent	Final concentration	Amount
bovine serum albumin frac V	1 g/L	100 mg
PBS	n/a	100 mL
Total	**n/a**	**100 mL**

Filter sterilize using a 0.22μm filter and store at 4°C for no longer than 3 days.

A.4   Step-by-step method details

A.4.1   Step 1A: Cell Induction

Timing: 26 hours

S RBD is expressed in yeast from a galactose-inducible promoter. The first step is to expand cells from a frozen stock and then induce S RBD display by growth on galactose.

1. Thaw 1mL stock of yeast cells harboring the S RBD display plasmid of the desired library for identifying escape mutants and WT for controls (Confirm that antibodies selected bind and inhibit the displayed WT RBD). It is convenient to start this growth step in the morning so that cytometer screening can be performed during daytime hours on the following day.

2. Pellet the cells by centrifuging 1 min at 16000 x g using a microcentrifuge, remove supernatant, and resuspend in SDCAA$^+$ (recipe). Transfer culture to a 14 mL culture tube for volumes lower than 1.5 mL (VWR Cat#10127-334) or a 250 mL shaker flask for large volumes (Fisher Cat# 10040F). Incubate cells at 30°C and 300rpm for 4-6 hours.

   Note: Do not grow cultures of more than 1.5 mL volume in 14 mL tubes, as they are insufficiently aerated for optimal display.

3. Measure the culture OD_{600} with a spectrophotometer. If $1.2 < OD_{600} < 5$, harvest the cells by centrifugation for 1 min at 16000 x g in a microcentrifuge and resuspend to $OD_{600}=1$ in $SGCAA^+$ (recipe).

4. Transfer the appropriate amount of the culture to a new 14mL culture tube or a 250mL shaker flask as necessary and incubate 22h at 22℃ and 300rpm in an incubator shaker. As previously stated, we recommend using 14 mL culture tubes for the volumes lower than 1.5 mL and 250 mL shaker flasks for larger volumes.

A.4.2   Step 1B: Competitive binding reactions

Timing: 4 hours

The steps here are very similar to the preparatory procedure. Thus, the steps that follow in this section are consolidated and shortened. For additional guidance and notes, please refer to Preparatory Work C.

CRITICAL: The binding reactions cannot be stored to sort another day. If multiple antibodies are being tested, we recommend splitting the samples into multiple days.

5. Transfer the culture to a 1.5 mL microcentrifuge tube for low volumes or a 15 mL conical tube for larger volumes and centrifuge for 1 min at 16000 x g using a centrifuge. Measure the culture OD600 and resuspend the cells in ice cold PBSF to an $OD_{600}=2$ for the controls and to an $OD_{600}=10$ for sorting. Keep this suspension on ice during the whole procedure except when otherwise specified. We typically perform a 10-fold dilution of culture to determine the OD_{600}. The yeast culture will not necessarily increase in OD_{600} during this overnight induction.

Screening Unlabelled Cells

6. In a 1.5 mL microcentrifuge tube, add 5 μL cells with 45 μL of PBSF ('unlabeled' cells).

Incubate 30 min at RT (20-25°C). Centrifuge 1 min at 16000 x g using a microcentrifuge and resuspend the cells in 100μL ice cold PBSF. Screen 25,000 unlabeled cells and set the FSC-A/FITC$^+$ gate as in **Figure A.2B**.

Check Display of S RBD

7. Label 5 μL cells with 44 μL of PBSF and 1 μL of anti-c-myc FITC in a 1.5mL microcentrifuge tube. Mix by pipetting up and down several times. Incubate the cells on ice for 10 min protected from light.

8. Centrifuge the cells 1 min at 16000 x g using a microcentrifuge. Wash twice with 200 μL ice cold PBSF and remove supernatant. Cell pellets can be stored on ice until ready to screen.

9. Resuspend the cells in 100μL ice cold PBSF and screen 25,000 cells. If the yeast does not display the S RBD, do not prepare the binding reactions and refer to Troubleshooting 1.

Check Unlabeled Cells, ACE2 Binding, and Competitive Inhibition

10. These are the recommended though not essential controls for WT and each library sorted. The unlabeled cells will be used to set the correct gates on FACS for each different library sorted. The ACE2 binding is the positive control and the competitive inhibition the negative control. Set up three separate labeling reactions in three different 1.5 mL microcentrifuge tubes using the cells at OD_{600}=2. Incubate 30 min at room temperature (20-25°C). Ensure that the biotin:protein molar ratio has been adjusted before starting the escape mutant identification, else, refer to 'Confirm that antibodies selected bind and inhibit the displayed WT RBD'.

	Tube 1	Tube2	Tube 3
Library Cells	5μL	5μL	-
WT Cells	-	-	5μL
Biotinylated ACE2	-	75nM	-
Free ACE2	-	-	10μg/mL
PBSF	**To 50μL total volume**	**To 50μL total volume**	**To 50μL total volume**
	Incubate for 30 minutes at room temperature (20-25°C).		
Biotinylated ACE2	-	-	75nM
	-	-	Incubate for 30 minutes at room temperature (20-25°C).

(a) Centrifuge 1 min at 16000 x g using a microcentrifuge. Wash twice with 200 μL PBSF and remove supernatant.

(b) Resuspend the second and third pellet with 50 μL fluorescent labelling mix. The tubes should be labeled with 0.6 μL anti-c-myc FITC, 0.25 μL SAPE and 49.15 μL PBSF. The first tube must not be labelled. Keep the fluorophores on ice and protected from light at all times.

(c) Incubate 10 min on ice and protect from light.

(d) Centrifuge 1 min at 16000 x g using a microcentrifuge. Wash twice with 200 μL PBSF and remove supernatant. The cell pellet can be stored on ice until ready to screen. We do not recommend storing on ice for longer than 1h as fluorophores might dissociate,

resulting in a lower signal.

(e) Resuspend the cells in 100 μL PBSF and screen 25,000 cells.

Competitive Binding Reactions to Identify Escape Mutants (Figure A.2A)

11. In a 1.5mL microcentrifuge tube add 150 µL of cells, PBSF, and 10 μg/mL antibody to a final volume of 400 μL. Incubate 30 min at RT (20-25°C). Then, add biotinylated ACE2 to a final concentration of 75nM. Again, incubate 30 min at RT (20-25°C).

   CRITICAL: If multiple samples are sorted, it is better to incubate pellets on ice before the next step. Fluorescent label right before sorting to ensure maximum fluorescence. To minimize binding loss, stagger the binding reactions to minimize the storage time so that when the reaction is completed, the sample can be sorted.

   (a) Centrifuge 1 min at 16000 x g using a microcentrifuge. Wash twice with 1 mL PBSF and remove supernatant. Cell pellets can be stored on ice for $\sim$4 h without significant binding loss if all the supernatant is removed.

   (b) Resuspend each cell pellet with 200 μL fluorescent labelling mix (1.2 μL anti-c-myc FITC, 1 μL SAPE and 196.6 μL PBSF per reaction plus an additional 10%). Keep the fluorescent labelling mix on ice and protected from light at all times.

      CRITICAL: Better results are obtained if the cells are fluorescently labelled right before sorting.

   (c) Incubate 10 min on ice and protect from light.

   (d) Centrifuge 1 min at 16000 x g using a microcentrifuge. Wash twice with 1 mL PBSF and remove supernatant. Cell pellets can be stored covered and on ice until ready to screen.

Optional: While sorting, the fluorescence signal might decrease over time. If the decrease is significant, divide the sample into two 500 μL samples and spin down again, remove supernatant and spin down to make sure no liquid is left in the sample. Store each pellet on ice until ready to use.

A.4.3   Step 1C: Fluorescent activated cell sorting (FACS)

Timing: approx. 40 min per sorted sample

Cells are sorted by FACS using four separate gates (**Figure A.2**). We first set an SSC/FSC gate for isolation of yeast cells. The second gate set is on FSC-H/FSC-A to discriminate single yeast cells from budding cells and/or aggregates. The third gate is an FSC-A/FITC$^+$ gate that selects the cells displaying the RBD on their surface and excludes cells not displaying RBD. The fourth gate is a square gate collecting the top 2% by a PE$^+$/FITC$^+$ (**Figure A.2B**). The first three gates are set using unlabeled cells as described below. The collection PE$^+$/FITC$^+$ gate is adjusted for each sorted sample to collect the cells giving the top 2% PE fluorescence signal. We set a 2% limit for collecting the population of the PE fluorescence signal for two reasons. First, collecting the entire population would require collecting considerably more cells to obtain the same quality in results since different PE signal increases correlate with varying affinity. In addition, collecting the top 2% allows us to set a false discovery rate (FDR) given that cells with higher signal are likely to contain escape mutants.

12. Using the unlabeled cells, set the FSC/SSC, FSC-H/FSC-A, FSC-A/FITC$^+$ gates as shown in **Figure A.2B**.

13. Using the same unlabeled cells, obtain the reference population by collecting at least 200,000 cells by sorting only for passing through the FSC/SSC gate. This reference population contains all the mutations in the library. It will be the reference to calculate the enrichment ratio for each mutation. For this and all other cell populations, collect cells into a 14mL

culture tube (VWR Cat#10127-334).

14. Sort the control population. Here we use cells that have not labelled with antibody or ACE2 but have been fluorescently labelled with SAPE and anti-c-myc FITC. Collect 200,000 cells that pass through FSC/SSC, FSC-H/FSC-A, FSC-A/FITC$^+$, and the PE$^+$/FITC$^+$ gates.

15. For every other sample, collect 200,000 cells that pass through FSC/SSC, FSC-H/FSC-A, FSC-A/FITC$^+$, and the PE$^+$/FITC$^+$ gates. The square PE$^+$/FITC$^+$ gates will have to be reset for each sample to ensure 2% collection. Using our Sony SH-800 cytometer, it typically takes 40min to collect 200,000 cells in the control and sample populations. This time results from a sorting rate of 5-10,000 cells per sec and maintaining a 70% sorting efficiency.

> **Note:** To avoid complexity bottlenecks, it is important that the number of collected cells is at least 150X the theoretical library size. For a library with NNK mutations on 56 and 63 mutations for library 1 and 2 respectively, 168,000 cells for library 1 and 189,000 cells for library 2 must be collected. Results are improved with more collected cells.

Step 1D: Cell recovery

Timing: 3 days

Cells are recovered and yeast stocks are prepared.

Day 1

16. Centrifuge collection tube containing the collected cells 5 min at 3200 x g using a centrifuge. Collected cells are smaller than typical yeast bulk population, and therefore a longer centrifugation time of 5 min is required.

17. Remove approx. 80% of the supernatant and resuspend in 1.5mL SDCAA$^+$ at 30°C overnight (16h) at 300rpm in an incubator shaker in the 14mL culture tubes.

   Note: 14mL culture tubes cannot hold more than 1.5mL. Larger volumes are not well aerated. To use a larger resuspension volume, grow the culture in a 25 mL shaker flask.

Day 2

18. Passage cells into 3 culture tubes using 1.5mL SDCAA$^+$ on each. Incubate at 30°c and 300rpm overnight (16h).

Day 3

19. Prepare 1mL cell stocks at OD_{600}=4 in yeast storage buffer at -80°C.

   (a) Centrifuge the cell culture 2 min at 3200 x g.

   (b) Remove the supernatant and resuspend the cell pellet in yeast storage buffer (recipe) at OD_{600}=4. Stocks can be stored for over a year at -80°C. In total, at least 1 mL cell stock at OD_{600}=4 is needed for deep sequencing preparation.

A.4.4   Step 2A: Deep sequencing library preparation

Timing: 1-2 days

Plasmid DNA from cells is extracted and amplicons prepared for Illumina sequencing (**Figure A.3**). The protocol used is adapted from Method B detailed in Kowalsky et al. (2015) [61].

Day 1: Yeast Miniprep

20. Perform a yeast miniprep.
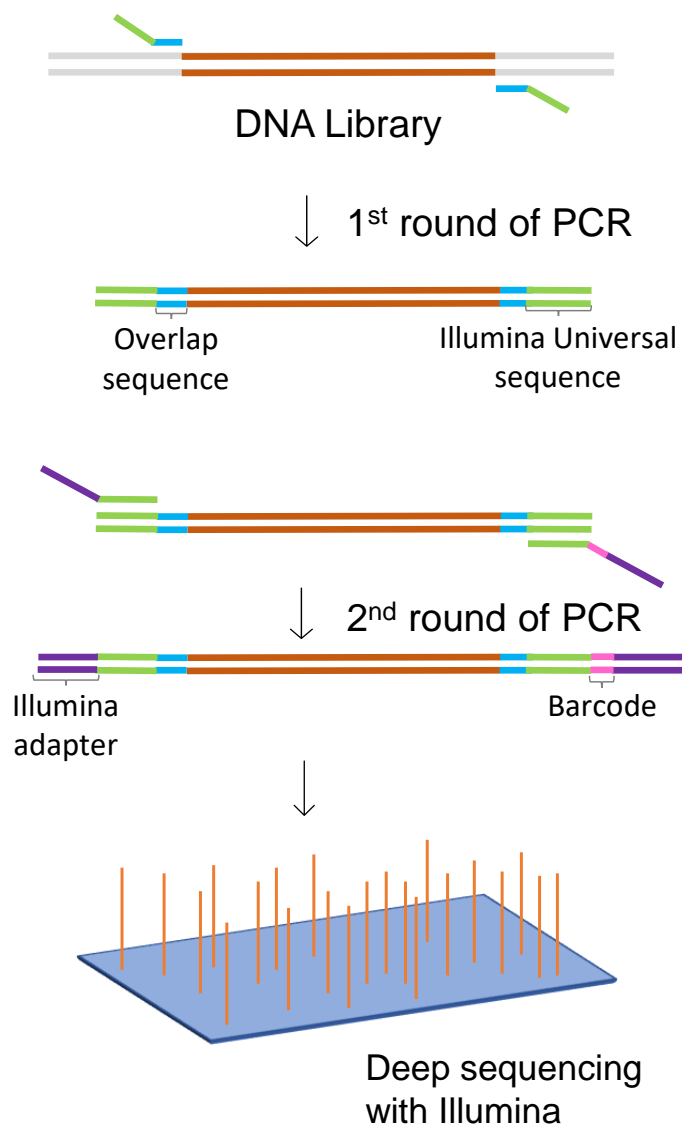
Figure A.3: **Schematic of deep sequencing library preparation.** Barcodes and Illumina adapters are added to the DNA library through two rounds of PCR. The amplicon is sequenced using an Illumina MiSeq

(a) On ice, thaw a cell stock prepared in step 23.

(b) Pellet the cells by centrifuging 1 min at 16000 x g using a microcentrifuge and resuspend in 200 μl Solution 1 from Zymo Yeast Plasmid Miniprep II kit.

(c) Add 5μl Zymolyase (5U/μl).

(d) Incubate in a 37°C incubator for 4h on an end-to-end mixer or incubate at 37°C with mixing by pipetting up and down (ten times) every hour.

(e) Perform 1 freeze-thaw cycle using a dry ice/EtOH bath followed by a 42°C incubation for 10 min. Alternatively, freeze at -80°C for 20 min and thaw at 42°C in a temperature bath.

(f) Add 200 μL Solution 2 from Zymo Yeast Plasmid Miniprep kit, mix end-over-end and let sit for 5 min. The solution will be transparent.

(g) Add 400 μL Solution 3 from Zymo Yeast Plasmid Miniprep kit mix end-over-end centrifuge 5 min at 16000 x g. The solution will turn yellow and precipitate will form.

(h) Transfer supernatant to a NEB miniprep column.

> **Note:** The NEB miniprep column has a higher DNA binding capacity (15 μg) compared to the Zymo miniprep columns (5 μg). The rationale for using NEB miniprep columns is that lysed cells contain large amounts of sheared yeast genomic DNA in addition to the desired plasmid DNA.

(i) Centrifuge 1 min at 16000 x g using a microcentrifuge.

(j) Add 700 μl Wash 1 buffer from NEB mini-prep kit and centrifuge 1 min at 16000 x g.

(k) Add 700 μl Wash 2 buffer from NEB mini-prep kit and centrifuge 1 min at 16000 x g.

(l) Add again 700 μl Wash 2 buffer from NEB mini-prep kit and centrifuge 1 min at 16000 x g.

(m) Decant the supernatant and centrifuge 1 min at 16000 x g to dry the column.

(n) Add 30 μl elution buffer from NEB mini-prep kit to elute the DNA and centrifuge 1

min at 16000 x g.

(o) Reload column with eluate and spin again.

21. Enzymatic cleanup of yeast genomic DNA. In PCR tubes add in order:

Lambda buffer 10X	2 μL
Miniprep DNA	15 μL
Exonuclease I	2 μL
Lambda exonuclease	1 μL
Final volume	**20 μL**

In a thermocycler

30°C	1:30 h
80°C	20 min
4°C hold	Forever

22. Plasmid cleanup. Using Monarch PCR & DNA cleanup kit according to the manufacturer instructions. Elute in 30 μL elution buffer.

Can divide into two days or continue same day from here.

Day 2: Sequencing Sample Preparation

23. The Illumina universal sequences are attached to the amplicon with an overlapping region included in the primers (**Figure A.3**) using a first PCR.

Primers used for first PCR (corresponding to "Overlap sequence" on **Figure A.3**):

Name	Description	Sequence	Section of RBD amplification	Library compatibility
IFU-104	L1-Inner-FWD	gttcagagttctacagtccgac gatctggaggaggctctgg	333-437	pJS699-L1, pIFU001-T1 and pIFU002-T1
IFU-105	L1-Inner-REV	ccttggcacccgagaattccac caagctataacgcagcc		
IFU-106	L2-Inner-FWD	gttcagagttctacagtccgac gatcggctgcgttatagcttgg	438-537	pJS699-L2, pIFU001-T2 and pIFU002-T2
IFU-107	L2-Inner-REV	ccttggcacccgagaattccag ccccctttgtttttaaccaa		

DNA (from PCR cleanup)	15 μL
Fwd inner primer (10μM)	2.5 μL
Rev inner primer (10μM)	2.5 μL
2X Q5 Master mix	25 μL
ddH$_2$O	5 μL
Final volume	**50 μL**

1st PCR cycling conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	98 °C	1 min	1
Denaturation	98 °C	10 sec	25 cycles
Annealing	64 °C	20 sec	
Extension	72 °C	1 min	
Final extension	72 °C	2 min	1
Hold	4 °C	forever	

24. 24. Perform a PCR cleanup to remove unwanted single stranded DNA. To the PCR products add:

PCR Products	50 μL
Exonuclease I	5 μL
rSAP	10 μL
Total volume	**65 μL**

In the thermocycler

37°C	20 min
80°C	20 min
4°C hold	Forever

25. Attach the barcodes and the Illumina adapter with a 2nd PCR reaction. The barcode must be unique for each sample to allow the analysis of different samples on a single MiSeq run. Primers used on 2nd PCR (corresponding to the Illumina primers – "Illumina adapter" and "Barcode" – in **Figure A.3**):

Description	Sequence
Illumina_FWD (RP1)	aatgatacggcgaccaccgagatctacacgttcagagttctacagtccga
Illumina_REV	caagcagaagacggcatacgagatnnnnnngtgactggagttccttggcacccg agaattcca

Note: nnnnnn refers to unique RPI Illumina barcode (full list: Illumina Adapter Sequences and Kowalsky et. al 2015 [61]). In this study, we used adapters #1-20 but any set of adapters can be used.

H$_2$O	18 μL
DNA (from cleaning step)	2 μL
Fwd outer primer (10μM)	2.5 μL
Rev outer primer (10μM)	2.5 μL
2X Q5 MM	25 μL
Final volume	**50 μL**

2nd PCR cycling conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	98 °C	1 min	1
Denaturation	98 °C	10 sec	
Annealing	64 °C	20 sec	25 cycles
Extension	72 °C	1 min	
Final extension	72 °C	2 min	1
Hold	4 °C	forever	

26. Gel extract the amplified DNA using Monarch Gel Extraction kit. Select the approx. 500bp band. A clear band should be seen with the correct size (515 bp for Tile 1 and 487bp for Tile 2) on the SYBR safe gel. If this band is not seen, refer to Troubleshooting 3. If products cannot be seen in a SYBR safe gel, there will not be enough DNA for sequencing (minimum 2ng DNA is required).

27. Clean up with Agencourt AMPure XP following standard procedure on a 96 well format (Instructions For Use).

28. Quantification with Quant-iT™ PicoGreen™ dsDNA Reagent following standard procedure (Quant-iT™PicoGreen™dsDNA Reagent and Kits).

    (a) Prepare Quant-iT™reagent.

        (i) Thaw Quant-iT™reagent at room temperature (20-25°C) covered in foil.

        (ii) Prepare a working solution by diluting the Quant-iT™reagent 200-fold in TE.

        CRITICAL: The working solution must be prepared on the same day of the quantification and stay protected from the light, preferably wrapped in aluminum foil.

(b) In a black 96-well plate (Product # 3792), prepare a 1:2 standard curve starting at 100ng/mL of Lambda DNA using the first row of the plate. Add 100 μL in each well and include a blank with 100 μL TE.

(c) In the same plate, add 1 μL of purified sample in 99 μL TE.

Note: If the sample's concentration is ¡ 40 ng/μL, up to 2.5 μL of sample can be used. On the contrary, if the concentration is too high, necessary dilutions in IDTE can be carried out.

(d) Add 100 μL Quant-iTTM working solution to each well and incubate for 5 min at room temperature (20-25°C) while covered in foil.

(e) Measure the fluorescence using a plate reader (ex: 480nm, em: 520 nm). Subtract the blank from all the samples and use the Lambda DNA serial dilutions to generate a standard curve of fluorescence vs DNA concentration. With the curve, determine the concentration of each sample. In our hands the concentrations are between 6 ng/μL and 60 ng/μL.

29. Pool individual barcoded amplicons into a single test tube. In a 1.5 mL tube, mix equivalent molar amounts of each sample to have the same number of reads and submit to a sequencing facility.

A.4.5   Step 2B: Illumina sequencing

Libraries can be sequenced on an Illumina MiSeq using the MiSeq V2-500 cycle kit. While we have used the University of Colorado sequencing core, there are commercial and academic cores that offer sequencing as a fee-for-service (NovoGene or Rush University Medical Center, for example). The libraries have been split into two tiles assuming the use of 250 bp paired end sequencing, but our software supports longer paired-end reads as well. One critical detail essential for sequencing is

dealing with the low nucleotide diversity of the RBD libraries. We circumvent this issue primarily by instructing the sequencing core to add PhiX in the pooled library, which is a crucial step for correct clustering of the sequences. PhiX is a well characterized adapter-ligated library that helps balance nucleotide diversity. We have obtained good sequencing results with 35% PhiX. Refer to Troubleshooting 4 if the DNA is not clustering correctly.

We typically obtain around 90% coverage for all the libraries screened.

A.4.6   Step 3: Run software and generate results

Timing: <1 hour

This is the primary computational step in the protocol. This software package can perform all necessary analysis and interpretation of results. With the completion of this step, potential escape mutations are identified for a given antibody.

Software description:

This software is divided into two separate modules. The first module manages the input FASTQ files and is abbreviated 'dms' for **d**eep **m**utational **s**canning. The second module, titled 'analysis', is responsible for interpreting the output generated from the dms module. Apart from specifying which experiment to analyze before the analysis module is run, the transition between these two modules is seamless.

This protocol results in a total of three output files. The dms module generates one of these files and the analysis module identifies the other two. The output from the dms module is a comma-separated values (CSV) file that identifies the single mutations, their corresponding counts in each population, and an enrichment ratio (ER) that compares these counts. Next, the analysis module generates a CSV file with additional columns to the file generated in the dms module. The new columns identify the following for each observed mutation: the threshold ER value determined from

the control data for a given tile, the p-value calculated by the one-sided exact Poisson rate ratio test, the FDR value, and the minimum number of nucleotides away the given mutation could be from the wild type sequence. The second output of the analysis module is a heatmap in Microsoft Excel that identifies potential escape mutations in blue. Refer to **Figure A.4** for details on the outputs mentioned above.
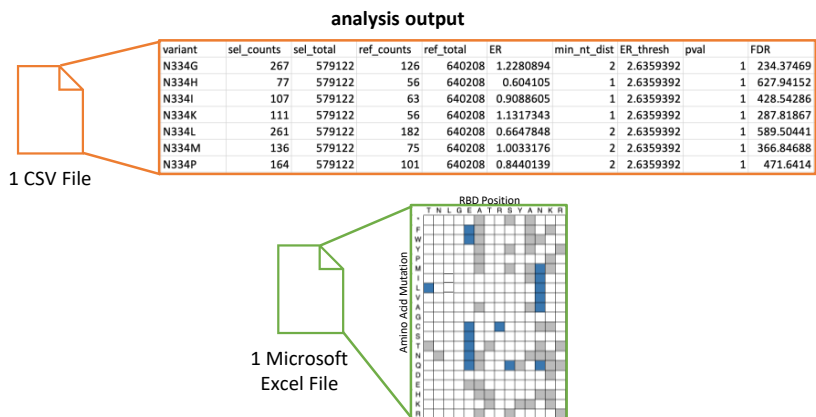
Figure A.4: **Schematic of deep sequencing library preparation.** The CSV file is first generated, adding the results of statistical calculations. The second output from the analysis module is the Microsoft Excel document containing a heatmap with escape mutant residues identified. In the heatmap, wild type residues and positions are listed in the top two rows of the file and each possible mutation is listed in the left column. Three colors exist in each Excel output file: grey cells indicate that the mutation was not observed in the population; white cells indicate mutations that were observed in the population but did not meet the criteria for escape mutant; and, blue cells represent mutations that meet the necessary criteria to be classified as potential escape mutants for the given antibody.

The dms module assumes that amplicons are of the same fixed length and that each amplicon is read in both the forward and reverse directions, leading to a pair of FASTQ files. When the module is run, it merges the paired reads of the FASTQ files, collapses identical synonymous mutations into single counts, and subsequently calculates enrichment ratios between specified populations (see 'Quantification and statistical analysis section'). A detailed explanation of this process has already been published [159, 160]. The analysis module performs the necessary analytical steps to identify the escape mutations. While a thorough explanation of the statistical approach is detailed in Quantification and statistical analysis, please refer to **Figure A.5** for an overview of the entire

process.

30. Complete the configuration file to include the details of the experiment. For the dms module, the user should fill out the configuration file to include all experiments. Therefore, the module can be run once and the analysis can be streamlined. An example configuration file is on Github: `https://github.com/WhiteheadGroup/SpikeRBDStabilization.git`. We recommend that this example file is copied and that the original document remains available for reference. For an explanation of each section and option in the configuration file, please refer to **Table A.3** and **Table A.4**.

   (a) If the experiment of interest only uses one tile, refer to Troubleshooting 5.

   CRITICAL: The example configuration file is already filled out so that users only need to complete the file names, the file paths, and create names for their experiments. If the original libraries are used, no additional changes are needed to the configuration file. All other parameters and details are set as default and do not need to be changed for a successful run of the software.

31. Run the software by execution of commands at the command line. Some command line flags can be used here to override information set in the configuration file. Details about these command line arguments are in **Table A.5**.

   (a) Run the dms module. This module can be run with following command (assume the configuration file is titled example.config and that it is found in the root directory): python3 -m dms –config example.config

   The dms module will deposit output files in a folder found in the root directory of the software unless the output directory is intentionally changed. The files are always deposited into the folder titled "Output". If that folder does not exist in the directory, an output folder will be created.

Note: The dms package is a general deep mutational scanning tool. It can be run independently of any other analysis and does not depend on the other module described here.

(b) Run the analysis module with the following command (again, assume the configuration file is titled example.config and that it is found in the root directory). The [Analysis] section of the configuration file must be updated before every run to represent accurate experiments. However, all other sections of the configuration file should remain the same for both modules: python3 -m analysis –config example.config

    (i) Details on the approach for this analysis can be found in the Quantification and statistical analysis section.

The analysis module will deposit output files in a folder found in the root directory. The files are always deposited into the folder titled "Processed". If that folder does not exist in the directory, the folder will be created.
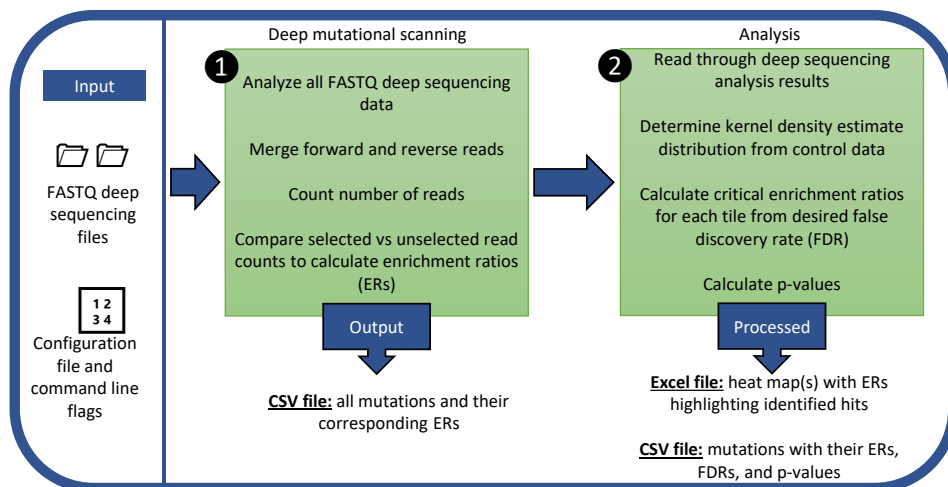
Figure A.5: **The flow structure of the software.** For each module (Deep mutational scanning; Analysis) the actions taken and output are listed.

Note: We provide the CSV outputs of the dms module on Github in the "Output" folder.

There are two files: one representing the control population (example_control.csv), cells without ACE2 labeling, and the other representing an antibody population for CC12.1 [3], titled example_CC12.1.csv. In addition, the example configuration file provided is already configured to properly run the analysis module with these example files. The only required change is to update the file paths to the Output folder containing the aforementioned example files. To run the analysis module with this data, follow the instructions in step 31b.

CRITICAL: The analysis module is written with the intention that it will be executed shortly after running the dms module. Therefore, if output files of the dms module are significantly edited, the analysis module may not run as expected

Table A.3: Descriptions of sections in the configuration file

Configuration file section	Description
Parameters	This section defines the basic parameters for the dms module. Available options (described in **Table A.4**): max_mismatches, min_quality, and fastq_file_dir.
Tile:Tx	This section defines a tile that will be analyzed. For every tile in an experiment, a new tile section should be created. To define a tile, change the name after the colon. Available options (described in **Table A.4**): wt_seq, first_aa, cds_start, cds_end, and positions. **Note:** this section should generally remain unchanged from the example configuration file. The option to change these parameters is provided as a convenience for anyone following these protocols with different libraries.

Table A.2 Continued

Configuration file setction	Description
Samples	This section defines the different samples in the experiment, connecting the tiles with the file names for the paired-end reads. All samples are in this section, no other [Samples] section is necessary to define multiple samples. The title for each sample is chosen by the user and should be entered without quotation marks. Each sample must be unique, and all samples should be options under [Samples]. The samples must be defined as follows: sample_name: **'Tx', 'FASTQ_filename1', 'FASTQ_filename2'** 'Tx' - the tile identifier that must match the tile name provided in the tile section. 'FASTQ_filename1' - the name of the FASTQ file that contains the sample's forward reads. 'FASTQ_filename2' - the name of the FASTQ file that contains the sample's reverse reads

Table A.2 Continued

Configuration file setction	Description
Experiments	This section defines each experiment by giving it a name and identifying the reference and selected populations. All experiments are in this section, no other [Experiments] section is necessary to define multiple experiments. The title for each experiment is chosen by the user and should be entered without quotation marks. Each experiment must be unique, and all experiments should be options under [Experiments]. The experiments must be defined as follows: **experiment_name: 'reference_sample', 'selected_sample'** 'reference_sample' - the name of the sample that contains the reference population 'selected_sample' - the name of the sample that contains the desired selected population ∗ Both 'reference_sample' and 'selected_sample' must be defined in the [Samples] section as options. The name of these samples must be copied exactly into quotes in the [Experiments] section.

Table A.2 Continued

Configuration file setction	Description
Proteins	This section combines different tiles that correspond to the same protein. This is so the information across all tiles for a given protein can be consolidated into single concise files.
	The title for each protein is chosen by the user and should be entered without quotation marks. Each protein must be unique, and all proteins should be defined under this section.
	The proteins must be defined as follows: **protein_name: 'experiment_name1', 'experiment_name2', . . . , 'experiment_nameN'**
	'experiment_name1' - the name of the experiment that is the first tile for the protein
	'experiment_name2' - the name of the experiment that is the second tile for the protein
	* All experiment_names must be defined in the [Experiments] section as options. The name of these samples must be compiled exactly into quotes in the [Proteins] section.
	It is possible to define a protein with a single experiment, but the single quoted experiment name must be followed by a comma in the definition.
Analysis	This section defines statistical parameters for the analysis module. The information provided here results in the final escape mutant hits identified. Available options (described in **Table A.3**): control_filepath, antibody_filepath, output_title, FDR, significance.

Table A.4: Descriptions of options for each section in the configuration file

Configuration file section	Configuration file options	Description
Parameters	max_mismatches	The maximum number of mismatches allowed when reading overlapping sequences of paired-end reads from FASTQ deep sequencing files. Reads with a higher number of mismatches are discarded. The input must be a positive integer or None. If None, then no reads will be discarded due to mismatches. The default value used in our analysis is 10.
Parameters	min_quality	The minimum sequencing quality score required to keep a read in the analysis. A read with a score lower than this minimum benchmark will be discarded. The input must be a positive integer or None. If None, then no reads will be discarded due to quality. The default value used in our analysis is 10.
Parameters	fastq_file_dir	This option allows the user to specify the directory that contains the FASTQ files to be analyzed. The file path should be a folder that contains FASTQ files with the file names provided in the configuration file. An absolute file path should be used. The input must be a string. An example for this value is: '/Users/lab/Documents/fastq_files'

Table A.3 Continued

Configuration file section	Configuration file options	Description
Tile:Tx	wt_seq	The nucleotide DNA sequence of the tile which includes the entire sequence of a read from deep sequencing. This generally includes the primers used. The input must be a string of capital letters. The default values used in our analysis is different for each Tile and shown below: Tile 1: 'TGGAGGAGGCTCTGGTGGAGGCGGTAGCG GAGGCGGAGGGTCGACAAACTTGTGCCCTTTTGG TGAAGTTTTTCAAGCCACCAGATTTGCATCTGTT TATGCTTGGAACAGGAAGAGAATCAGCAACTGTG TTGCTGATTATTCTGTCCTATATAATTCCGCATC ATTTTCCACTTTTAAGTGTTATGGAGTGTCTCCT ACTAAATTAAATGATCTCTGCTTTACTAATGTCT ATGCAGATTCATTTGTAATTAGAGGTGATGAAGT CAGACAAATCGCTCCAGGGCAAACTGGAAAGATT GCTGATTATAATTATAAATTACCAGATGATTTTA CAGGCTGCGTTATAGCTTGG'

Table A.3 Continued

Configuration file section	Configuration file options	Description
		Tile 2: 'GGCTGCGTTATAGCTTGGAATTCTAACAAT CTTGATTCTAAGGTTGGTGGTAATTATAATTACC TGTATAGATTGTTTAGGAAGTCTAATCTCAAACC TTTTGAGAGAGATATTTCAACTGAAATCTATCAG GCCGGTAGCACACCTTGTAATGGTGTTGAAGGTT TTAATTGTTACTTTCCTTTACAATCATATGGTTTC CAACCCACTAATGGTGTTGGTTACCAACCATACA GAGTAGTAGTACTTTCTTTTGAACTTCTACATGC ACCAGCAACTGTTTGTGGACCTAAAAAGTCTACT AATTTGGTTAAAAACAAAGGGGGC'
Tile:Tx	first_aa	The residue number of the first amino acid in the protein sequence. The defaults used in our analysis are: Tile 1: 333 Tile 2: 437
Tile:Tx	cds_start	The nucleotide position in a merged read is where the coding sequence starts (after primers or any other miscellaneous nucleotides at the start of the read). **Note:** this index is given using Python slicing conventions, so it is zero-indexed. For example, the sequence 'GATC' is numbered 0123. If the sequence starts at 'A', cds_start would be 1. The input must be a nonnegative integer. The default used in our analysis is different for each Tile and shown below: Tile 1: 43 Tile 2: 18

Table A.3 Continued

Configuration file section	Configuration file options	Description
Tile:Tx	cds_end	The nucleotide position in a merged read is where the coding sequence ends (before primers or any other miscellaneous nucleotides at the end of the read). **Note:** this index is given using Python slicing conventions, and so is the index of the first nucleotide that is not part of the coding sequence. For example, if the CDS is 'AT' in 'GATC', cds_end would be 3. The input must be a nonnegative integer. The default values used in our analysis is different for each Tile and shown below: <u>Tile 1:</u> 355 <u>Tile 2:</u> 321

Table A.3 Continued

Configuration file section	Configuration file options	Description
Tile:Tx	positions	The amino acid positions of interest in the experiment. If the user only desires certain positions in the output files, this is where those positions should be specified.
		The input must be a list of integers. For example: [22, 42, 44, 51]. If you desire only one residue, enter the integer as a list (e.g. [31]).
		The default positions used in our analysis is different for each Tile and shown below: <u>Tile 1:</u> 333, 334, 335, 339, 340, 344, 345, 346, 349, 351, 352, 354, 356, 357, 358, 359, 360, 362, 363, 364, 366, 367, 370, 372, 373, 375, 376, 377, 378, 380, 381, 382, 383, 384, 385, 386, 388, 389, 390, 394, 396, 405, 408, 411, 412, 413, 414, 415, 417, 420, 421, 424, 426, 427, 428, 430
		<u>Tile 2:</u> 437, 439, 440, 441, 444, 445, 446, 447, 448, 449, 450, 452, 455, 456, 457, 458, 459, 460, 462, 463, 464, 465, 466, 468, 469, 470, 471, 473, 474, 475, 477, 478, 479, 481, 482, 483, 484, 485, 486, 487, 489, 490, 492, 493, 494, 498, 499, 500, 501, 503, 504, 505, 506, 508, 516, 517, 518, 519, 520, 521, 522, 523, 527

Table A.3 Continued

Configuration file section	Configuration file options	Description
Analysis	control_filepath	The file path, including the file name, to the output file from the initial deep mutational scanning module for the control experiment. Absolute file paths should be used. Include extensions, if applicable. An example of this value is: '/Users/lab/Documents/rbd/Output/example_control.csv'. Analysis antibody_filepath The file path, including the file name, to the output from the initial deep mutational scanning module for the desired antibody experiment. Absolute file paths should be used. Include extensions, if applicable. An example of this value is: '/Users/lab/Documents/rbd/Output/example_CC12.1.csv'.
Analysis	output_title	The prefix to be used in the processed file names at the end of running the analysis module. An example of this value is: 'CC12-1'
Analysis	FDR	The target false discovery rate (FDR) for identifying an enrichment ratio threshold. This threshold is then used to determine escape mutant hits. The default used in our analysis is 1.
Analysis	significance	The p-value cutoff used to determine if a given enrichment ratio is significantly greater than the threshold value. The default used in our analysis is 0.01.

Table A.5: Command line flags that can be used when run-

ning dms module

Comand line flag	Description
'–use_multiprocessing' [boolean]	This flag provides an option for the user to use the multiprocessing module or to turn it off. The multiprocessing module reads FASTQ files using multiple cores and can speed up the run time. The default is set to True and the argument parser accepts yes, no, true, or false as options.
'–fastq_file_dir' [file path]	This flag allows the user to specify the directory that contains the FASTQ files to be analyzed. The file path should be a folder that contains FASTQ files with the file names provided in the configuration file. The default is an empty string.
'–output_dir' [file path]	This flag allows the user to specify the output directory where all CSV and Microsoft Excel files will be saved when the run is complete. The default is an empty string.
'–max_mismatches' [number]	The maximum number of mismatches allowed when reading overlapping sequences of paired-end reads from FASTQ deep sequencing files. Reads with a higher number of mismatches are discarded. The default is set to None which sets no limit to the number of mismatches and does not discard reads due to mismatches. Any other input should be a positive integer.

Table A.4 Continued

Command line flag	Description
'–min_quality' [number]	The minimum deep sequencing quality score required to keep a read in the analysis. A read with a score lower than this minimum benchmark will be discarded. The default is set to None which sets no lower limit on the quality of a read and accepts all scores. Any alternative input must be a positive integer.
'–min_ref_counts' [number]	The cutoff used to discard variants that have a low number of counts in the reference population. Variants with reference counts below this cutoff will be discarded. The default is set to 5 and inputs must be a nonnegative integer.
'–pseudocount' [number]	The pseudocount used when a variant contains counts observed in the reference population but not in the selected population. This is required because observing no counts in the selected population while having counts in the reference population will lead to an undefined enrichment ratio. The default is set to 1 and inputs must be a positive integer.

Expected outcomes

There are several expected outcomes from the experimental procedure. These outcomes are included in the protocol, but they are also written here for completeness. First, in the preparation steps, co-transformation into yeast should result in $>10^4$ transformants for the wild type S RBD (Preparatory Work B(ii), #3e) and $>1.2 \times 10^5$ transformants for each S RBD library [74] (Preparatory Work B(iii), #4j). Second, when making yeast stocks for the wild type S RBD (Preparatory Work B(ii), #3l), each 1.5 mL *S. cerevisiae* EBY100 culture will produce an average of six 1 mL

yeast stocks. Third, the FACS step sorting 200,000 cells from the competing assay (Step 1C, #15) should take around 40 min to collect if using our Sony SH-800 system. This time results from a sorting rate of 5-10,000 cells per sec and maintaining a 70% sorting efficiency. Finally, gel extraction in the sample preparation stage before sequencing (Step 2A, #26) should always give one clear band with the correct size on the SYBR safe gel (515 bp for Tile 1 and 487bp for Tile 2). Any other expected outcomes during the experimental stages are very small and listed with their associated steps.

After the experimental work, the major outcome of the computational step is a statistical measure of how likely each single point mutant present in the library is of being an escape mutant for a specific neutralizing antibody. This information is captured in a comma-separated values (CSV) file generated by the analysis module. The most important information in this CSV file is a false discovery rate (FDR) for each mutation. The FDR for a mutation quantifies how many mutations we would expect to have similar, or more significant, enrichment ratios and depths of coverage by random chance. Other information contained in the CSV file for the set of mutants includes their corresponding counts in each population, an enrichment ratio (ER) that compares these counts, a p-value calculated by the one-sided exact Poisson rate ratio test, and the minimum number of nucleotides away the given mutation could be from the wild type sequence. The second output of this software is a heatmap in Microsoft Excel that identifies potential escape mutations in blue. Refer to **Figure A.4** for details on the outputs mentioned above.

Quantification and statistical analysis

All statistical analysis is performed by the custom software developed for this protocol. This occurs entirely in Step 3 from the procedures above. To determine escape mutations, the antibody population must be compared against the control group. A control was used to help identify which mutations are randomly expected to appear in the given library. With this data, we can determine the mutations present in the selected antibody population that occur more frequently

than would typically be expected when looking at the control data. This information is captured as the FDR value. If these specific mutations have enough counts to be confident of their presence in the antibody library, then we call those variants escape mutations. A more detailed and formal explanation of the statistical approach follows.

First, the critical metric for determining escape mutants in this analysis is the enrichment ratio at each position. An enrichment ratio is the log2 transform of the frequency of an observed mutation in the selected population divided by the frequency of that mutation in the reference population (**Figure A.6A**). This calculation is performed by the dms module prior to running any statistical analysis.

When running the analysis module, a kernel density estimate (KDE) is first fit to the control data. This provides an empirical ER distribution for cells sorted without antibody or ACE2 labeling, which serves as a null hypothesis in which observed ERs are due to stochasticity in sorting and other non-relevant differences between variants. In determining which distribution to use for this data, we initially fit common probability distributions (**Figure A.6B**). However, KDE consistently provided a better fit to the data. In addition, a KDE eliminates the need to make the underlying assumptions common with parametric functions, yet still provides an accurate distribution [161]. Thus, given the unpredictable nature of the control data, a nonparametric probability distribution such as the KDE was chosen to ensure a reliable distribution for each given control group without making any assumptions about the distribution of data.

Using this KDE from the null distribution, a threshold value for enrichment ratios can be determined. We estimate the false discovery rate (FDR) for a given threshold ER_t as the number ERs we expect to find above that threshold in the null experiment. This is given by $FDR = (1 - F(ER_t)) \times N$, where F is the experimentally determined cumulative distribution function (CDF) of the control experiment ERs and N is the number of variants in the library. Rearranging, we find $ER_t = F^{-1}(1 - \frac{FDR}{N})$, where F^{-1} is the inverse CDF. For a library size using these

A

Mutation N501Y in Control		
	Selected Counts (S)	Reference Counts (R)
N501Y (j)	857	639
Total (T)	718708	660627

$$ER_j = \log_2\left(\frac{S_j/S_T}{R_j/R_T}\right) = \log_2\left(\frac{S_j R_T}{R_j S_T}\right) = 0.302$$

B

C
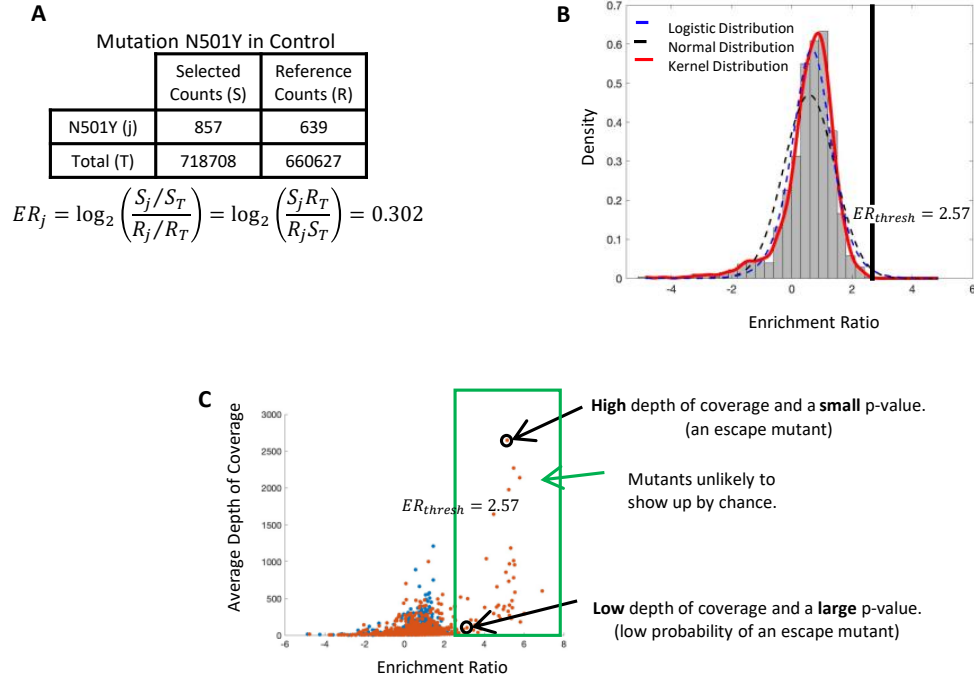
Figure A.6: **Quantification and statistical analysis. A.** An example of an enrichment ratio calculation. This calculation uses the total counts in the selected population, S_T, the total counts in the reference population, R_T, and the total counts of the given N501Y mutation in the selected and reference populations, S_j and R_j, respectively. **B.** Demonstrating that the best fit for this data is a kernel density estimate (KDE). The KDE fits very closely to the data, while other common probability distributions, such as the normal and logistic curves, do not. **C.** Showing the final statistical step that considers the depth of coverage when deciding to identify an escape mutation. For mutations with high depth of coverage, the returned p-value will be much smaller than for mutations with lower depths of coverage.

experiments (N=1120 for tile 1 or N=1260 for tile 2) and with an FDR=1, the typical range for this threshold is $2.2 < ER_t < 2.6$.

We then compare each observed ER in an experiment to this threshold to see which variants have enrichment ratios larger than ER_t. However, comparing just the observed ER to the threshold would ignore the depth of coverage for each mutation. Since the depth of coverage provides information about our confidence in the observed ER, we use it directly in the comparison. This is explained in detail below.

For a given experiment, we have two sets of reads $\{k_i^p\}$ and two totals n^p, where $p \in \{ref, sel\}$

is the reference or selected population. Note that, since we discard reads that don't encode single mutations, it is not generally true that $\sum_i k_i^p = n^p$. We approximate each read count k_i^p as being distributed as $k_i^p \sim Pois(r_i^p n_i^p)$ for some unknown rate r_i^p. If we knew this rate, the true enrichment ratio would be given by $ER_i = log_2 \left(\frac{r_i^{sel}}{r_i^{ref}} \right)$. Therefore, testing the hypothesis that $ER_i \geqslant ER_t$ amounts to testing that $\frac{r_i^{sel}}{r_i^{ref}} \geqslant 2^{ER_t}$, which we do with $k_i^{ref}, n^{ref}, k_i^{sel}$, and n^{sel} using a one-sided Poisson rate ratio test implemented in the Python package statsmodels [103]. The significance level of the test, α, is determined by the user and can be changed in the configuration file. Our analysis used $\alpha = 0.01$. Please refer to **Figure A.6C** for additional details.

Limitations

The method described has several experimental parts that introduce limitations to the protocol. First, this protocol identifies only the subset of neutralizing antibodies that directly compete with ACE2 for binding on S RBD. Neutralizing antibodies for SARS-CoV-2 can bind on Spike at locations other than the RBD [40, 39, 52], and some neutralizing antibodies recognize the N343 glycan that we have removed from the displayed S RBD to avoid hyper-mannosylation [72]. A different approach taken by the Bloom research group can be considered if non-competing antibodies must be analyzed [31, 66] that also bind on the S RBD. We have previously tried to display the S ectodomain in yeast but were not successful [3].

Second, this procedure is only designed to identify escape mutants resulting from single amino acid mutations. The smaller single mutation library limits our ability to test the possibility that a combination of amino acid mutations could work in conjunction and ultimately lead to escaping antibody binding. Larger and more complex libraries would be needed to test this type of escape mutant. In addition, the custom software developed with this protocol is not written to identify such scenarios.

Third, the residues examined in this study are all located on the surface of S RBD. While

ACE2 and other antibodies will only interact directly with these surface residues, it is possible that buried residues could have indirect impacts on binding. Furthermore, since we test monomeric S RBD some surface mutations identified as putative escapes by this method could produce steric clashes in the full trimeric S. Each mutation is pleiotropic with unpredictable potential changes in the fitness of the virus. Thus, the escape mutants identified by this protocol will necessarily be a subset of the escape mutants for these same positions observed in nature.

Finally, given that a mutation was observed in the population, the output is binary (yes/no for an escape mutant). Some true escape mutants may be missed if the mutation does not appear with high enough frequency in the antibody population. This is especially true if ER values are close to the cutoff, if the mutation is not sampled in the reference library, or if the control population has some aberrations that skew the ER threshold higher.

A.5   Troubleshooting

Problem 1:

S RBD does not display on the yeast surface or has a very low display. This problem may rise in Before you begin step 14 or Step-by-step method details step 8.

Potential solution:

Low display percentages occur with insufficient culture aeration during growth. Ensure the 14 mL culture tube is shaking and that the culture volume is 1.5 mL or less. Induction time and temperature were optimized for our laboratory settings and may need to be altered to optimize display. Other common temperatures are 18°C and 30°C.

If necessary, a 0.2 w/v% galactose can be added to the SDCAA$^+$ and 0.2 w/v% dextrose can be added to the SGCAA$^+$. If the RBD is still not displaying, co-transform again.

Problem 2:

No or low PE signal increase when binding biotinylated ACE2 to the displayed RBD. This problem may arise in Before you begin step 20 or Step-by-step method details step 10e Tube 2.

Potential solution:

Increase the biotin to protein ratio. If the signal increases as the ratio increases, use the necessary ratio to see a good PE signal increase.

Problem 3:

Product from the second PCR when preparing the libraries for deep sequencing does not show on a SYBR safe gel in Step-by-step method details step 26.

Potential solution:

First check that there is some product on the right band size using the more sensitive SYBR gold stain. If so, rerun the PCR steps using more cycles. Else, cells could have been contaminated during cell recovery or the yeast miniprep did not have a high enough yield.

Problem 4:

Illumina deep sequencing does not cluster correctly. This would occur after preparing for sequencing in Step-by-step method details Step 2B.

Potential solution:

Increase PhiX percentage to increase nucleotide diversity to promote correct clustering.

Problem5:

The example configuration file is not set up by default for running with only one tile. Thus, using the software for an experiment that only has a single tile may result in an error if the configuration file is not formatted correctly through the initial setup in Step-by-step method details step 30.

Potential solution:

To run a single tile, continue to add the experiment in the [Proteins] section. However, after identifying the first tile, add a comma and then a space (leaving the rest of the line blank). For example, if the experiment name is '6-29' and the chosen title for the protein is '1_6-29' then use the following format in the [Proteins] section.

Control_1: '1_Control',

This should fix any issues.

A.6 Resource availability

A.6.1 Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Tim Whitehead timothy.whitehead@colorado.edu.

A.6.2 Materials availability

All plasmids and mutational libraries used in this work are available from AddGene (`https://www.addgene.org/Timothy_Whitehead/`).

Data and code availability All scripts used to process and analyze deep sequencing data are freely available on Github (`https://github.com/WhiteheadGroup/SpikeRBDStabilization.`

`git`).

A.7   Acknowledgments

A.8   Author contributions

Developed computational software: CMH, PJS; Developed experimental protocol: IMFU, PJS, TAW; Wrote the protocol: CMH, IMFU, PJS, TAW.

Appendix B

User-defined library generation by Golden Gate assembly

B.1   Background & protocol overview

Golden Gate assembly is a one-pot DNA assembly technique that uses temperature cycling with a type IIS restriction enzyme (BsaI-HFv2 reported here), T4 ligase, and input DNA to clone multiple DNA fragments sequentially into a vector. Generally, for this protocol, a PCR thermocycler cycles between a 37°C step and a 16°C step. During the 37°C step BsaI generates four base pair overhangs downstream of its GGTCTC(X) binding motif and during the 16°C step the reannealed DNA is ligated by T4 ligase. BsaI sites on the DNA fragments are designed such that overhangs will match complementary overhangs on upstream and downstream fragments or vector ends, allowing the fragments to be assembled in order. Additionally, the orientation of the BsaI sites on the vector and fragments are such that they are removed from the final product. Thus, cycling of these cutting and ligating steps continues to act on the input DNA and leaves the final product untouched, resulting in high incorporation efficiencies.

Here we detail a version of this protocol that can be applied to any vector and protein coding sequence to generate large, site-specific, combinatorial libraries by assembling synthetic mutagenic DNA fragments that we denote cassettes into a destination vector. Cassettes are linear dsDNA and can be purchased as synthetic dsDNA (e.g. gBlocks/eBlocks from IDT) or assembled from oligonucleotide or oligo pool ssDNA. Destination vectors have some, or all, of the protein coding

sequence replaced with a selective marker, such that mutated cassettes coding for that sequence replace the marker during the Golden Gate reaction, resulting in functional plasmids. Users can follow our protocol to make their own destination vector or use one of our premade vectors; our protocol provides a detailed explanation for designing cassettes that work with pND003, pND004, and pND005. After the destination vector is obtained, cassette(s) (one to four demonstrated) are generated from ssDNA by PCR and co-incubated with the destination vector in a Golden Gate reaction to assemble the library, which is then transformed into E. coli. We describe a streamlined protocol such that, once the destination vector is in hand, new libraries can be built in a single day.

The protocol is listed in three sections: (1.) Design and construction of the destination vector; (2.) Design and construction of double stranded cassette DNA; (3.) Golden Gate assembly and transformation into E. coli. Additionally, we provide examples for different steps based on the creation of the SARS-CoV-2 S RBD (333-541) yeast display libraries (RBD libraries).

B.2   Materials and supplies

∗ All customized DNA fragments were purchased from IDT

Item	Supplier	Catalog number
Section 1: Construction of destination vector		
pETconNK	Klesmith et al. 2017[162]	`https://www.addgene.org/81169/`
knock-out gene	Lee et al. 2015[130]	`https://www.addgene.org/65154/`
Section 2: Construction of dsDNA cassette		

Materials and supplies table Continued

Item	Supplier	Catalog number
Cassette 1 ultramer/opool	This study	Table S4
Cassette 2 ultramer/opool	This study	Table S4
Cassette 3 ultramer/opool	This study	Table S4
Forward primer	This study	Table S5
Q5 High-Fidelity 2X Master Mix	NEB	M0492S
KAPA HiFi Ready Mix	Roche/Fisher	7958927001
Nuclease Free Water	IDT	11-05-01-14
Ultrapure Agarose	Invitrogen	16500-500
TAE Buffer (Tris-acetate-EDTA) (50X)	Thermo Scientific	B49
SYBR Safe DNA gel stain	Invitrogen	S33102
Gel loading dye, Purple	NEB	B7024
Nucleo-Spin® Gel and PCR Clean-up kit	Macherey-Nagel	740609.250
Section 3: Golden Gate assembly and transformation into E. coli		
Destination vector	This study	Section B.3.1
dsDNA Cassette 1	This study	Section B.3.2
dsDNA Cassette 2	This study	Section B.3.2
dsDNA Cassette 3	This study	Section B.3.2
BsaI-HF v2	NEB	R3733S
T4 DNA ligase	NEB	M0202S
T4 DNA ligase reaction buffer	NEB	B0202S
Nuclease Free Water	IDT	11-05-01-14
XL1-Blue Competent Cells	Agilent	200249
TransforMax EC100 Competent Cells	Biosearch Technologies	EC10010
Corning® Square Bioassay dish	Sigma	CLS431272

Materials and supplies table Continued

Item	Supplier	Catalog number
Glass Beads	Sigma	G8772
Kanamycin	GoldBio	K-120-25
ZymoPURE II Plasmid Midiprep kit	Zymo Research	D4201

B.3 Protocol

B.3.1 Section 1: Design and construction of the destination vector

Background: We built three destination vectors (MBP E. coli expression, yeast surface display, yeast two-hybrid), each containing a GFP marker and BsaI recognition sequences next to high-fidelity overhangs (**Figure B.1**). If any of these destination vectors are suitable, section 1 can be skipped.

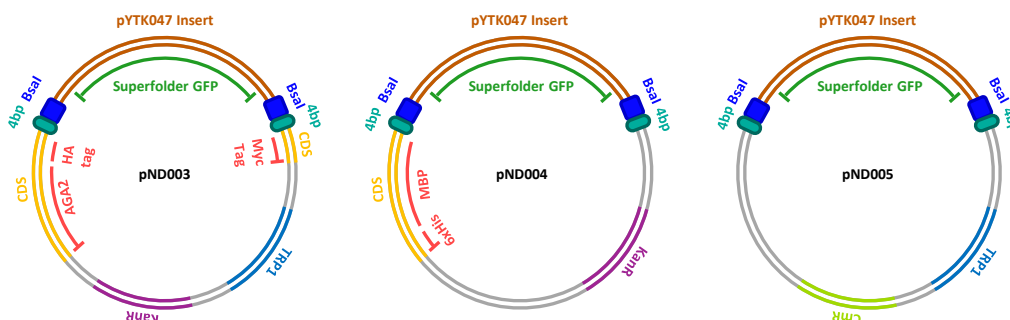Figure B.1: Premade general-use destination vectors for Golden Gate-based cassette mutagenesis

If these destination vectors are unsuitable, you can create your own destination vector. The following describes how to make a destination vector using Gibson Assembly, although a variety of cloning techniques could be used.

1: Pick a backbone

Pick a backbone for your destination vector which contains the protein of interest and the

desired vector components (tags, promoters, selection markers, etc). Check to see if the vector already contains BsaI sites. If so, these should be removed by site-directed mutagenesis before proceeding.

2: Choose a selection marker

Our lab has used both GFP and RFP but other selection markers, such as LacZ[127, 87], LacI[87], β-galactosidase[87] or sacB[134], have been demonstrated to be effective for GG assembly.

3:Choose the start and stop points for your cassettes

Before picking the start and end points for your library, you must choose the general region covered by your cassettes. Here you have two options. If you already have your protein in the desired vector, your cassettes can cover only the coding sequence containing your mutation sites. This allows for the fewest number of cassettes, is the simplest to design, and is ideal for large protein-coding sequences but requires that a new destination vector be created if you mutate a different region of your protein. Alternatively, you can have your cassettes begin and end outside the protein coding region, such that any library can be made for any protein without building a new destination vector (as with our premade destination vectors).

Once you have chosen the region covered by your cassettes, pick a set of high-fidelity 4 base pair overhangs from Table 1 in Potapov et al[87]. Then find a four base pair sequence near the beginning and end of that region from that set. These will be the start and end points for your cassettes.

4: Design primers for creation of the destination vector

Design primers to clone the desired selective marker and BsaI sites into the destination vector, in place of the backbone sequence covered by the cassettes. If using Gibson Assembly, this includes two primers to amplify the selection marker from its source as well as two primers to amplify

the backbone vector. The primer set for amplifying the backbone should add BsaI sites directly adjacent to the first and last overhangs and 20 base pairs of homology with the selection marker from each end of the PCR fragment. Importantly, the BsaI sites should be in between the overhang and selection marker sequence (**Figure B.2**).

5: Create the destination vector

If using Gibson Assembly, the template vector and selection marker insert should be amplified using a standard PCR reaction. The resulting PCR products are then gel purified and combined by Gibson Assembly[141].

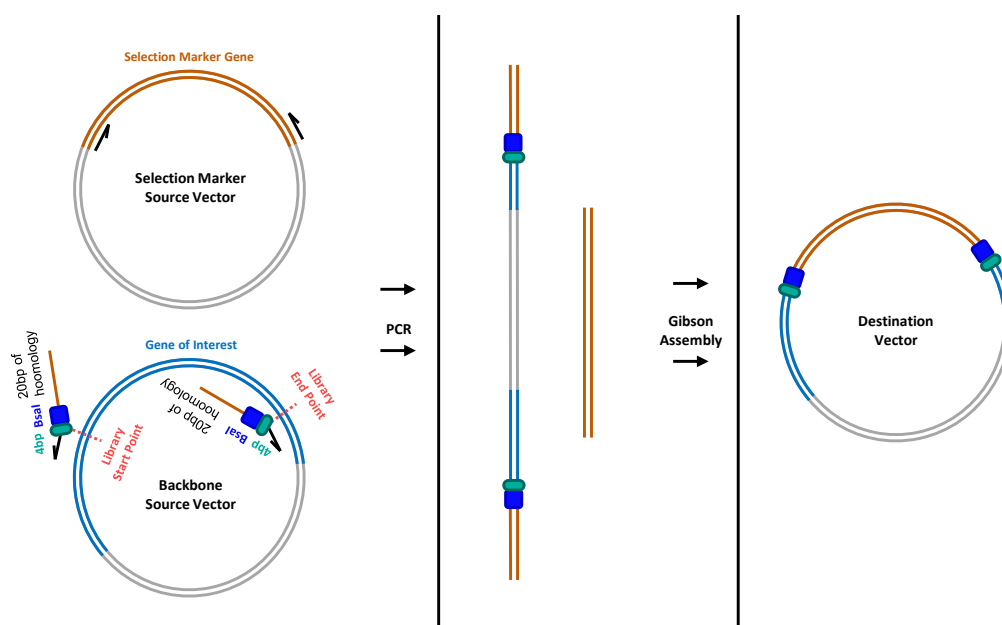Figure B.2: Building a destination vector using Gibson Assembly

B.3.2   Section 2: Design and construction of cassettes

Background: After designing a destination vector or choosing to use one of our premade vectors, you must design cassettes that contain your mutated protein-coding sequence, high-fidelity overhangs, and a BsaI recognition sequence. Additional sequences must be included when designing

cassettes for use with pND003, pND004, or pND005 (described below). These cassettes are ordered as synthetic ssDNA fragments and are converted to dsDNA using a reverse primer.

1: Cassette designs

The first step in designing cassettes is choosing your library's start and end points with respect to your protein coding sequence. If you created your own destination vector, as per section 1, these have already been chosen. If you are using our premade vectors, these can be chosen arbitrarily. Next, you must decide the number and length of your cassettes, which is usually based on the maximum length of the synthetic DNA fragments you are using (It is important to remember that the length of each fragment must include at least 7 base pairs on either side of the coding sequencing for the BsaI site). This will then determine the general areas in your coding sequence where each cassette will begin and end.

Within these areas, look for four base pair sequences that match high-fidelity overhangs from the set you used for the destination vector or that are listed for our premade vectors at the bottom of Table B.2. Note that all overhangs must be from the same set, cannot be used more than once and that residues coded for by base pairs in the overhangs cannot be mutated. These overhang sequences will be the starting and ending points of each cassette. If designed correctly, the first and last four protein-coding base pairs of each cassette will match the first or last four protein-coding base pairs of the adjacent cassettes (see Example B.4). Next, you will add BsaI recognition sequences directly adjacent to the overhangs (**Figure B.3**). Note, it is a good idea to check that any mutations you make do not introduce a BsaI site into your coding sequence. We also added short sequences outside of the BsaI site, if no other sequence was present, to ensure BsaI binding; we have not tested whether this is necessary to achieve high transformation efficiencies.

2: Cassettes designs for use with pND003, pND004, and pND005

Skip if not using our premade destination vectors.

Short sequence we included to support binding of BsaI to its recognition sequence. Removing it may not affect efficiencies but has not tested by us.	**GCCGT** ───	Only needed when working with our premade destination vectors or designing destination vectors with cassettes that extend beyond the protein coding sequence. Constant sequences that code for promoters, stop codons, terminators, and tags in source vectors. Exact sequences that match our premade vectors are found in **Supplemental Table 3** (colored in red).	**NNNNNNNNNN** ───
BsaI recognition sequences (forward/ reverse).	**GGTCTCA/TGAGACC** ▬▬▬	User's protein coding sequence. X represents the first and last mutation sites in the cassette	**NNXN........NXNN** ▬▬▬ ⋯⋯ ▬▬
4 base pair high-fidelity overhang sequence. When working with our premade destination vectors, the initial overhang in the first fragment and final overhang in the last fragment must contain a predetermined overhang that matches the vector. Exact sequences are found in **Supplemental Note, Table 1** (colored in cyan).	**NNNN** ⬭	Optional sequence that can be added if there are not enough bases after the last mutation site to design a reverse primer with an appropriate Tm	**NNNN** ───

Generalized DNA Fragment Designs
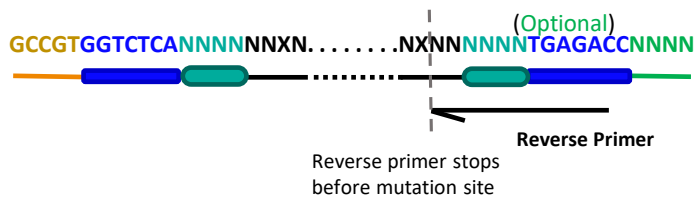
GCCGT**GGTCTCA**NNNNNNXN........NXNNNNNN**TGAGACC**NNNN

(Optional)

Reverse Primer

Reverse primer stops before mutation site

First DNA Fragment for Premade Vectors

(Specific to premade vector)

(Optional)

GCCGT**GGTCTCA**NNNN**NNNNNNNNNN**NNXN........NXNNNNNN**TGAGACC**NNNN

Reverse Primer

Last DNA Fragment for Premade Vectors

(Specific to premade vector)

(Optional)

GCCGT**GGTCTCA**NNNNNNXN........NXNN**NNNNNNNNNN**NNNN**TGAGACC**NNNN

Reverse Primer
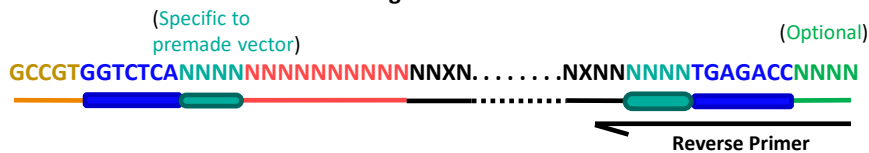
Figure B.3: Generalized design schematic of cassettes for Golden Gate-based cassette mutagenesis

Cassettes that will be used with pND003, pND004, or pND005 should be designed as described above, with two important exceptions:

First, cassettes should use overhang sequences as outlined in Table B.2. All our destination vectors were made with high-fidelity overhangs from Set 1 from the Table D.7 in Potapov et al4. Once you have chosen a destination vector, the initial overhang from your first cassette and the final overhang from your last cassette should be the ones listed in the sequence for the chosen vector, as seen in Table B.2. The rest of your overhangs should come from the remaining sequences in Set 1, found at the bottom of Table B.2.

Second, your first and last cassettes should include additional sequences, specific to the chosen destination vector, that are shown in red in Table B.2. These code for promoters, terminators, and continuations of tag sequences leading up to the overhang sequences. Note that stop codons do not need to be added to your sequence as these are already coded for in each vector.

Table B.2: Sequence specific designs of cassettes for use with pND003, pND004, and pND005. Coloring of bases corresponds to schematics in **Figure B.3**.

	First Cassette	**Last Cassette**
pND003	GCCGTGGTCTCACGGTAGCGG AGGCGGAGGGTCGGCTAGCCA TNNNN...NNNNNNNNTGAGAC CNNNN	GCCGTGGTCTCANNNNNNN N...NNNNCTCGAGGGGGGCGGA TCCGAATGAGACCNNNN
pND004	GCCGTGGTCTCACGGTCGTCAG ACTGTCGATGAAGCCCTGAAAG ACGCGCAGACTGGCGGCGGGGG AGGTNNNN...NNNNNNNNTGAG ACCNNNN	GCCGTGGTCTCANNNNNNN N...NNNNTGAGATCCGGCTGCT AACAAAGCCCGAAAGGATGAGA CCNNNN
pND005	GCCGTGGTCTCATCCTGAAA GNNNN...NNNNNNNNTGAGAC CNNNN	GCCGTGGTCTCANNNNNNN N...NNNNTGAGTCGTATTACCT CAGCCAAGCTAATTCTGAGAC CNNNN

High-fidelity overhangs that can be used in conjunction with pND003, pND004, and pND005	TGCC, GCAA, ACTA, TTAC, CAGA, TGTG, GAGC, AGGA, ATTC, CGAA, ATAG, AAGG, AACT, AAAA, ACCG

3: Design primers for creation of double stranded cassette DNA

Next, a single reverse primer must be designed to create double stranded cassette DNA. The primer should not overlap with any mutation site. If mutation sites are located close to the overhang sequences, additional bases might need to be added to the end of a cassette to have a long enough complementary sequence for primer binding (Supplemental Figure 3). In this study, we used primers with melting temperatures (Tm) ranging from 56°C to 72°C with comparable results. For the RBD library, we used NEB Tm calculator to determine the primer melting temperature and the IDT Oligoanalyzer™tool to analyze the possible hetero-dimers and the Gibbs free energy of the designed primers. We designed primers to avoid large negative ΔG of any possible mismatch (homodimer, unintended heterodimer, secondary structure) limiting this to >-22 kcal/mol. Additionally, we designed the $\Delta\Delta$G of the intended heterodimer to any other structure to be <-20 kcal/mol. Inspection of the types of secondary structures may also be helpful to look for sequences which can amplify in an unintended manner from an annealed 3' end (e.g. a hairpin or dimer that anneals the 3' end with an extended 5' end can serve as a template for amplifying products like "primer dimers").

4. Generate cassette dsDNA from synthetic ssDNA.

Our experimental protocol below starts at this step, in which dsDNA is generated from single-stranded synthetic DNA fragments. We found that smearing sometimes occurred but that using at least 35 PCR cycles leads to sufficient amplification of the desired product over background. The PCR products are then gel purified and can be stored until the next step.

Day 1

1. Dilute the synthetic DNA fragments and the reverse primers to a final concentration of 10 μM in Nuclease Free Water. (See Example B.4 for equations used to resuspend the synthetic DNA)

Note: The IDT Resuspension Calculator tool can be used.

2. Generate dsDNA cassettes

(a) Assemble the PCR reactions as follows:

Cassette (10 μM)	1.25 μl
Primer reverse (10 μM)	1.25 μl
Q5 2X Master Mix	12.5 μl
Nuclease Free Water	10 μl
Reaction volume	**25 μl**

Note: We have used KAPA HiFi Ready Mix (Roche) and Phusion polymerase (NEB) with comparable results.

PCR cycling conditions			
Steps	**Temperature**	**Time**	**Cycles**
Initial Denaturation	98°C	30 s	1
Denaturation	98°C	10 s	
Annealing	56°C*	30s	35 cycles
Extension	72°C	1:30 min	
Final extension	72°C	2 min	1
Hold	4°C	hold	

∗**Note:** We have demonstrated generation of dsDNA using primers with a Tm of 56°C as well as 72°C.

(b) Purify linear DNA by gel electrophoresis and gel extraction. One clear band is expected for each fragment. Purify all excised fragments using a Nucleo-Spin® Gel and PCR Clean-up kit (Macherey-Nagel). Elute in 20μl Nuclease free water. (See Example B.4; if trouble getting the right size band see Troubleshooting B.5)

(c) Quantify the linear DNA products by measuring A$_{260}$ using a spectrophotometer. We often obtain concentrations of approx. 40 ng/μl.

B.3.3   Section 3: Golden Gate assembly and transformation into *E. Coli*

Background. Use a Golden Gate reaction to assemble the dsDNA fragments into your destination vector. This section takes about 5 hours, including bacterial transformation, and can be combined with section 2 for a single day library generation protocol

Day 1 (Cont.)

1. Assemble the pieces with a Golden Gate reaction.

   (a) Per 40 fmol of input DNA, assemble the reaction as follows, adding the enzymes last. (See Example B.4 for the calculations used)

dsDNA fragments	40 fmol each
Destination vector	40 fmol
10x T4 DNA ligase buffer	2.5 μl
BsaI-HF-v2	1 μl (20 units)
T4 DNA ligase	1 μl (400 units)
Nuclease Free Water	up to 25μl
Reaction volume	**25** μl

   Note: input DNA or total reaction size can be scaled up linearly, as shown in the main text.

   Note: NEBridge Golden Gate Assembly kit (NEB) can also be used with comparable results following manufacturer's instructions.

   (b) Set the reaction on a thermal cycler using the conditions below:

Thermal cycler conditions		
Temperature	**Time**	**Cycles**
37°C	1 min	60 cycles
16°C	1 min	
37°C	5 min	1
65°C	10 min	1
4°C	infinity and beyond	

(c) To concentrate the DNA, perform a PCR clean-up using a Monarch® PCR & DNA clean-up kit (NEB), eluting in 6 μl of nuclease free water.

2. Transform into *E. coli.*

(a) Transform the 6 μl of eluent into electrocompetent cells. (See Example B.4)

   (i) Optionally, transform 0.1 ng of pUC18 (XL1 Blue) or pUC19 (TransforMax) into electrocompetent cells to assess competency.

(b) Recover the cells in 1 ml SOC or TB media for 1 h at 37°C with shaking at ∼300 rpm.

(c) Do 6 serial dilutions of the recovered cells and plate 10 μl of each dilution on an LB agar plate, containing the appropriate antibiotic, to assess transformation efficiencies. Plate the remainder of the cells on a LB agar square Bioassay dish, containing the appropriate antibiotic, using plating beads. Incubate the plates at 37°C overnight.

Day 2

3. Check the transformation efficiency and if it is above the desired efficiency, scrape the Square Bioassay dish. Add ∼3 ml of LB to the plate, using an L shaped cell scraper or a bended glass pasteur pipette, scrape the cells off the agar plate. Collect the cells in a corner and transfer them to a 50ml falcon tube by pipetting. Repeat this process five to

seven times for a total 15-20 ml cell volume or until the plate is clear of cells. (See Example B.4 to calculate the desired transformation efficiency)

Note: Depending on the destination vector, it might take up to two days to see green colonies. A good control to compare fluorescence of the uncut destination vector is to transform the neat destination vector by itself.

4. Make *E. Coli* glycerol stocks by spinning down 1.5 ml of the scraped cells and resuspending in 700 μl LB media and 300 mul 50 w/v% glycerol.

Note: Since the desired transformation efficiency is 100 fold above the library size, 1.5ml of the scraped cells will have a good representation of the library.

5. Prepare 10 mL of the scraped cells using a ZymoPURE Midiprep kit (Zymo Research) to obtain the plasmid library.

B.4   Examples

Example 1:

In the case of the RBD library, positions 333-416 and 510-541 were not mutated, therefore, they are part of the destination vector. For the RBD we wanted to mutate separately the epitopes of class 1 and class 2 antibodies7. Therefore, we divided positions 400 to 509 into 3 cassettes: cassette 1 containing positions 400-436, cassette 2 containing positions 436-472 and cassette 3 containing positions 472-509. This means that positions 400, 436, 472 and 509 can not be mutated (Figure S2).

Example 2: If the ultramer ordered contains 4 nmol of DNA, resuspending in 200 μl will give a 20μM concentration.

$$4nmol \times \frac{1\mu mol}{10^3 nmol} \times \frac{l}{20\mu mol} \times \frac{10^6 \mu l}{l} = 200\mu l$$

The oligo pool ordered contained 10pmol of DNA per individual oligo with 1644 oligos. Since we want the final concentration to be per pool:

$$10pmol \times 1644oligos \times \frac{1nmol}{10^3pmol} = 16.44nmol$$

Resuspending in 164.4 μl will give a 100μM concentration of the pool.

$$16.44nmol \times \frac{1\mu mol}{10^3nmol} \times \frac{l}{100\mu mol} \times \frac{10^6\mu l}{1l} = 164.4\mu l$$

Example 3: The fragments for the RBD are $\sim$170 bp, so the 170bp band was excised and gel extracted.

Example 4:

$\ast$ All dilutions are in nuclease free water.

Cassette	Length (bp)	Concentration from PCR (ng/μl)
1	167	33.56
2	168	40.24
3	172	30.31

Cassette 1

$$167bp \times \frac{650Da}{1bp} \times \frac{1g}{1Da \cdot mol} \times 40 \cdot 10^{-15}mol \times \frac{10^9ng}{1g} = 4.34ng$$

$$4.34ng \times \frac{1\mu l}{33.56ng} = 0.13 \rightarrow 1.3\mu l 1 : 10dilution$$

Cassette 2

$$168bp \times \frac{650Da}{1bp} \times \frac{1g}{1Da \cdot mol} \times 40 \cdot 10^{-15}mol \times \frac{10^9ng}{1g} = 4.37ng$$

$$4.37ng \times \frac{1\mu l}{40.24ng} = 0.11 \rightarrow 1.1\mu l \ 1:10 dilution$$

Cassette 3

$$172bp \times \frac{650Da}{1bp} \times \frac{1g}{1Da \cdot mol} \times 40 \cdot 10^{-15} mol \times \frac{10^9 ng}{1g} = 4.47ng$$

$$4.47ng \times \frac{1\mu l}{30.31ng} = 0.11 \rightarrow 1.1\mu l \ 1:10 dilution$$

Cassette 1	1.3 μl
Cassette 2	1.1 μl
Cassette 3	1.1 μl
Destination vector	2.3 μl
T4 DNA ligase buffer	2.5 μl
BsaI-HF-v2	1 μl
T4 DNA ligase	1 μl
Nuclease Free Water	14.7 μl
Reaction volume	**25** μl

Example 5: XL1-Blue Competent Cells (Agilent) or TransforMax EC100 Competent Cells (Biosearch Technologies) have been used with comparable results. In our hands when transforming 0.1ng of pUC18 into XL1-Blue we obtain $1.0e^5$ transformants ($1.0e^9$ transformants/μg DNA) and when transforming 0.1ng of pUC19 into TransforMax EC100 we obtain $6.2e^5$ ($6.2e^9$ transformants/μg DNA).

Example 6: The RBD libraries have a theoretical library size on the order of 10k protein-encoding variants. To ensure that all the variants are transformed, we want a 100-fold coverage. Therefore, the minimum desired transformation efficiency is $1e^6$.

B.5   Troubleshooting

Troubleshooting 1: If a single band is not observed, increase the extension time and/or increase the number of cycles. If Q5 1X Master Mix (NEB) does not give the desired results, the reaction can be set up with KAPA HiFi Ready Mix (Roche). Alternatively, using "touchdown PCR"8 may also minimize smearing.
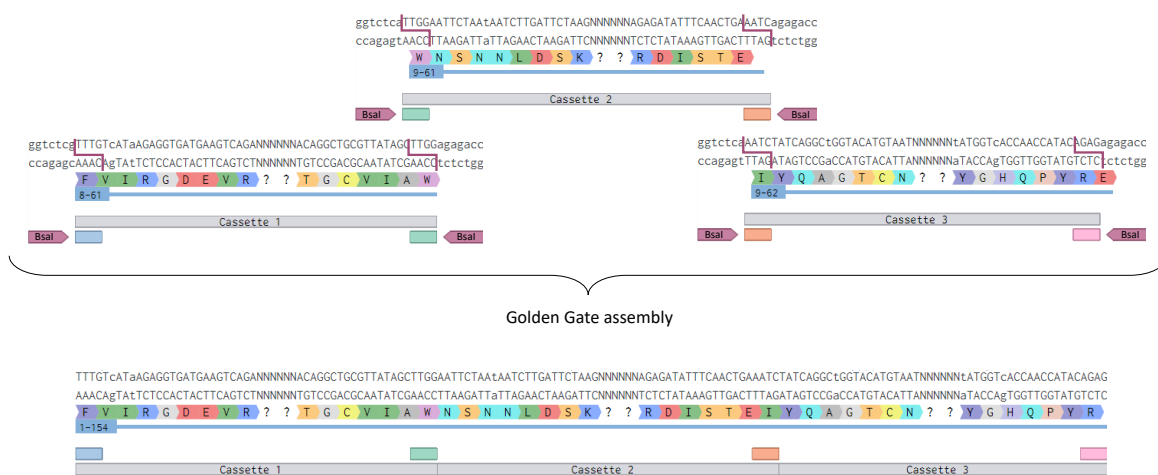
Figure B.4: Schematic of cassette assembly and location of the BsaI overhangs. The BsaI overhang at the end of cassette 1 matches the beginning of cassette 2 (green). Since those 4 bp allow the assembly, the codon(s) containing the overhang nucleotides cannot be mutated. Notice that the BsaI sites are located on the outside of each cassette and are not part of the assembled product.

Appendix C

Supplementary Figures
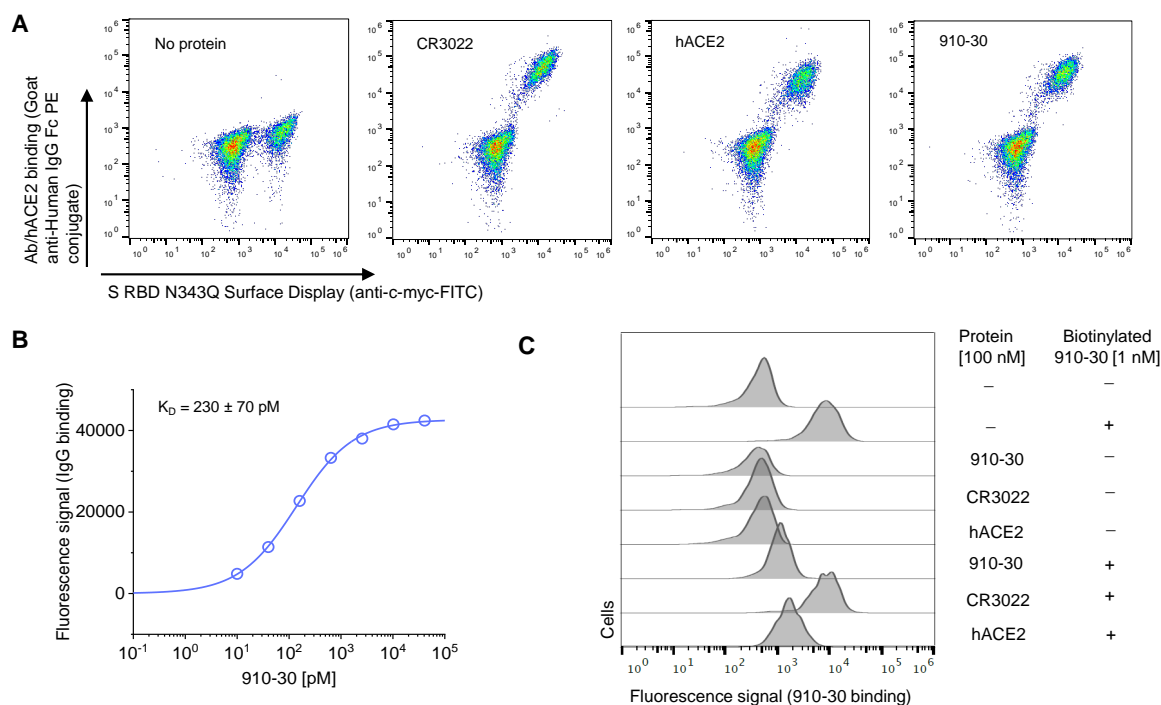
Figure C.1: **910-30 recognizes RBD in a glycan-independent manner and competes with hACE2 for binding.** **A.** S RBD N343Q with a C-terminal myc epitope tag was displayed on the surface of yeast and labeled with no protein or 1 nM of CR3022, human ACE2-Fc (hACE2), or 910-30. Cells were washed, secondarily labeled with anti-c-myc-FITC and Goat anti-Human IgG Fc PE conjugate, and read on a Sony SH800 cell sorter. Biological replicates were performed on two different days. **B.** Yeast cell surface titrations of 910-30 IgG against aglycosylated S RBD yield a KD of 230 ± 70 pM. Technical triplicates were performed for two biological replicates (n = 6), and error reported is 2 s.e.m. **C.** Competition binding experiments of free 910-30, hACE2 and CR3022 with biotinylated 910-30. Technical triplicates were performed for two biological replicates (n = 6).
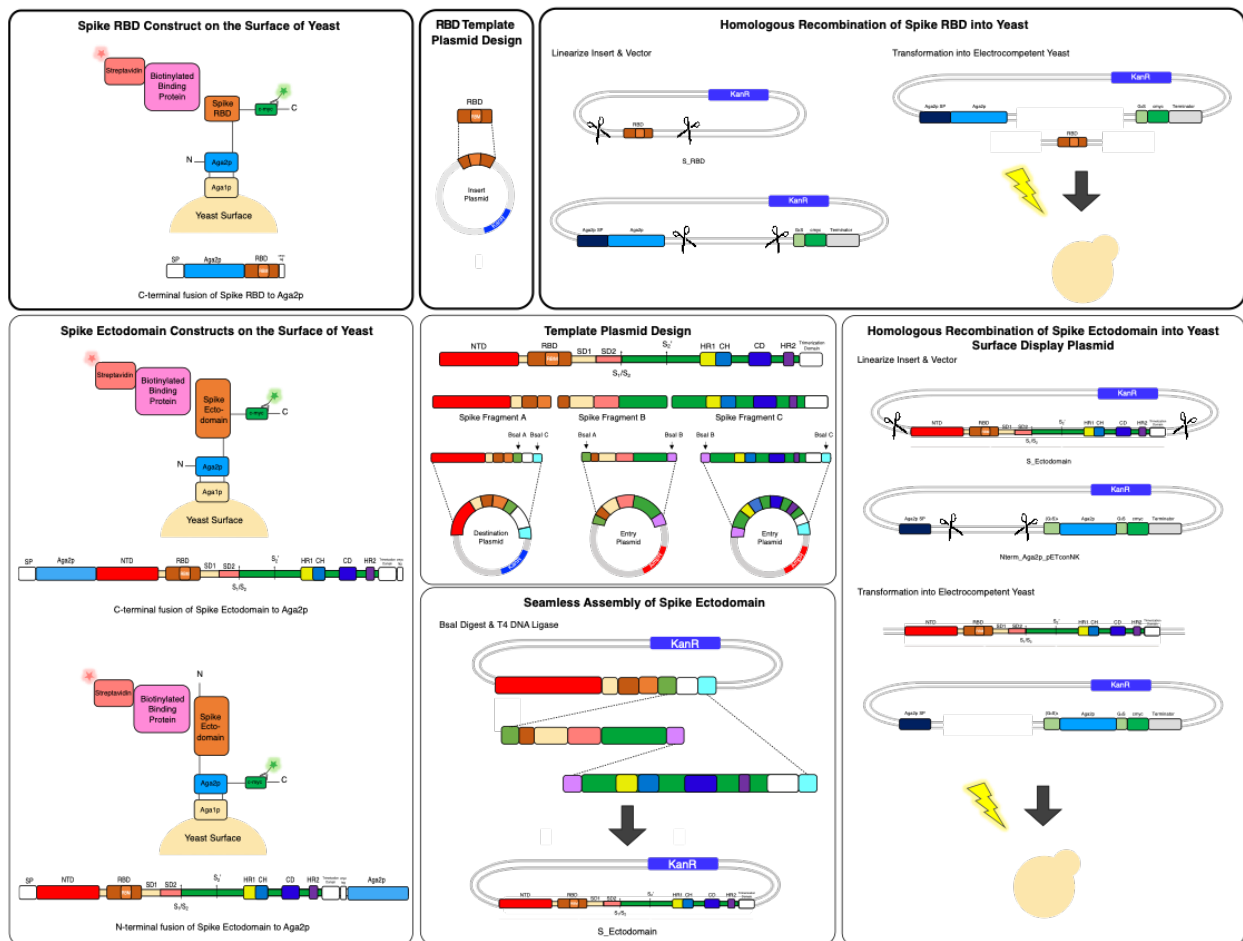
Figure C.2: **Overview of yeast display constructs used in screening S RBD and prefusion stabilized S ectodomain libraries.** Related to STAR Methods.
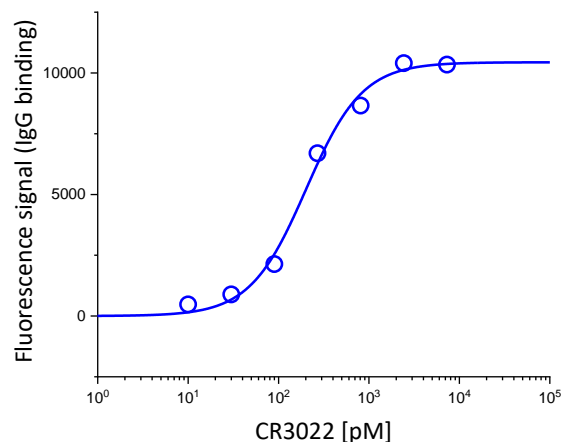
Figure C.3: **Non-neutralizing antibody CR3022 dissociation constant compared to literature.** Yeast cell surface titrations of non neutralizing CR3022 IgG against aglycosylated S RBD yield an apparent K_D of 171 ± 25 pM. Technical triplicates were performed ($n = 3$), and error reported is 2 s.e.m. Data for the HKU 910-30 nAb is from Banach et al, 2021. [28, 2]. Related to **Figure 3.1**.
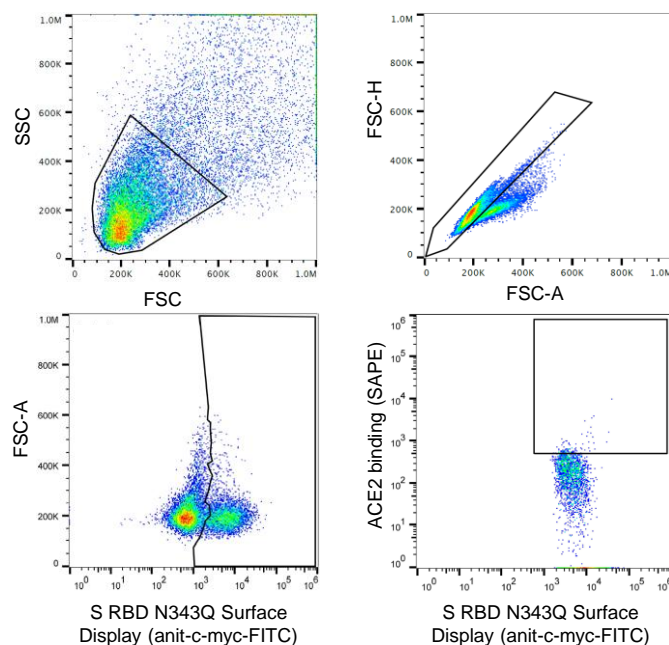
Figure C.4: **FACS sorting gates used to collect S RBD escape mutants.** Representative sorting gates used for all escape mutant FACS screens. The three gates were SSC/FSC; FSC-H/FSC-A to discriminate single yeast cells; FSC-A/FITC$^+$ to select cells displaying the RBD on their surface; and SAPE$^+$/FITC$^+$ to identify mutants that allow ACE2 binding in the presence of 10 μg/mL antibody. Shown here are gates for the antibody CC6.29. Related to **Figure 3.1**.
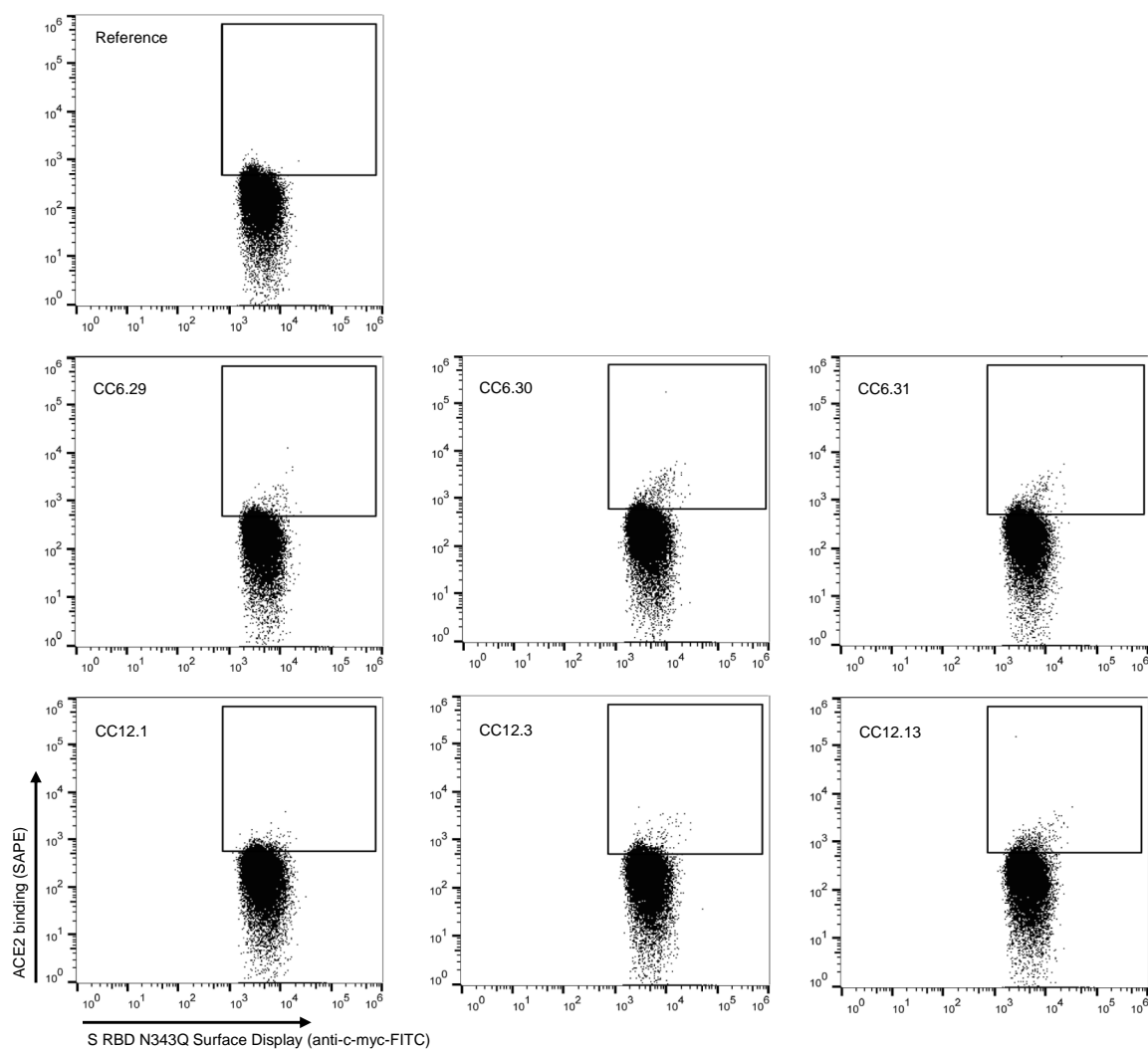
Figure C.5: **Sorting gates for competitive binding experiments.** PE/FITC cytograms for aglycosylated S RBD yeast libraries sorted using competitive binding with ACE2-Fc. The top cytogram shows the control experiment with no ACE2 labeling. Gates represent the top 2% of the FITC$^+$ cells by PE signal for each antibody used in the study. Related to **Figure 3.1**.
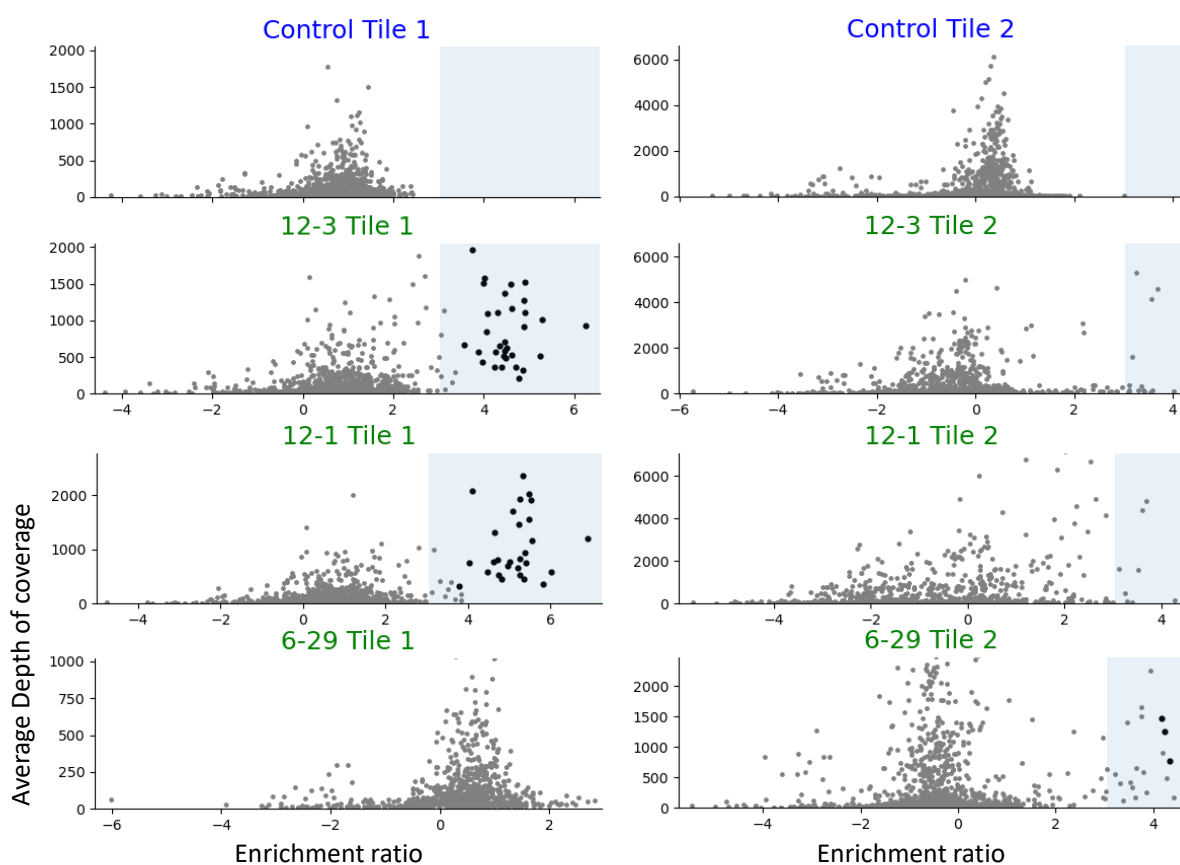
Figure C.6: **Per-mutation enrichment ratio (ER) distributions as a function of average depth of coverage** control (top) and CC12.3, CC12.1, CC6.29 nAb competing experiment (bottom). ER thresholds determined for FDR = 1 are shown in the control panels (top). Hits (with ER greater than the threshold at p ≤ 0.01) are shown with larger black dots. Related to **Figure 3.1**.

199
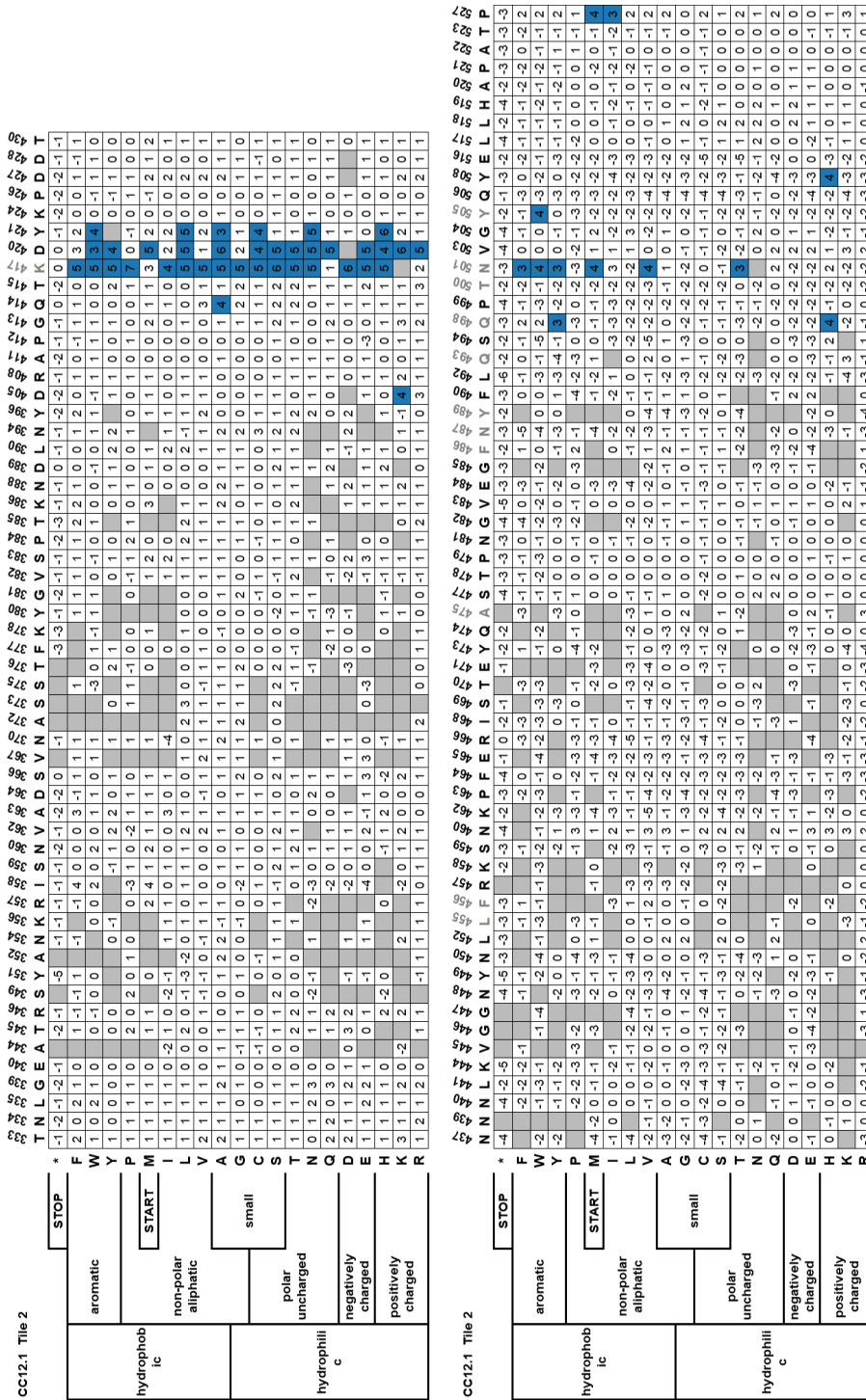
Figure C.7: **CC12.1 heatmap** using FDR <0.1 as cut-off. Blue - escape mutant hit, white – not a hit, grey – mutation not present. Position numbers in grey indicate ACE2 footprint. Related to **Figure 3.2**.

200

Figure C.8: **CC12.3 heatmap** using FDR <0.1 as cut-off. Blue - escape mutant hit, white – not a hit, grey – mutation not present. Position numbers in grey indicate ACE2 footprint. Related to **Figure 3.2**.
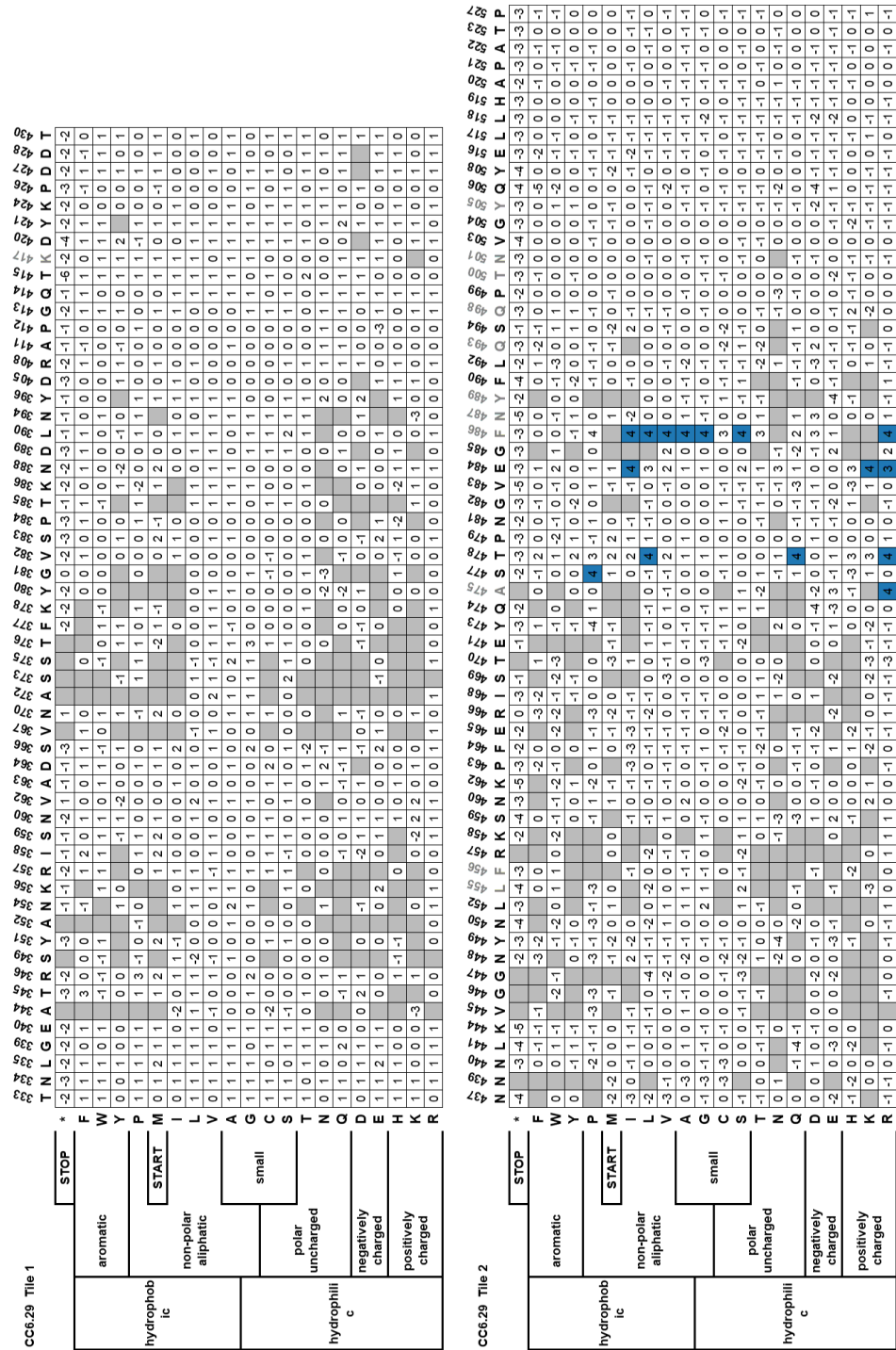
Figure C.9: **CC6.29 heatmap** using FDR <0.1 as cut-off. Blue - escape mutant hit, white – not a hit, grey – mutation not present. Position numbers in grey indicate ACE2 footprint. Related to **Figure 3.2**.
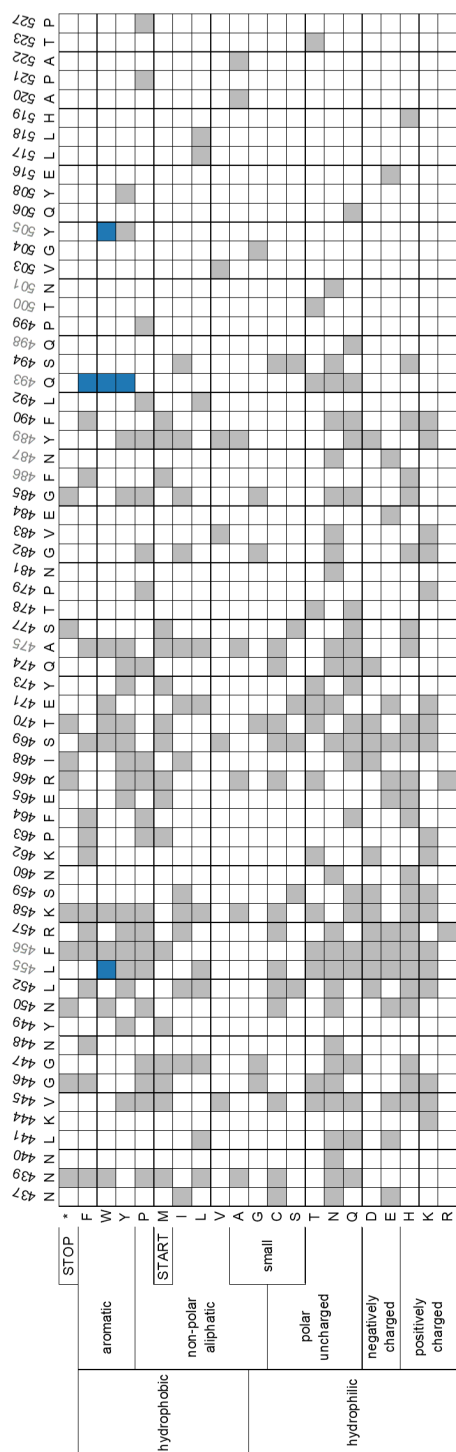
Figure C.10: **CC6.31 heatmap** using FDR <0.1 as cut-off. Blue - escape mutant hit, white – not a hit, grey – mutation not present. Position numbers in grey indicate ACE2 footprint. Related to **Figure 3.2**.
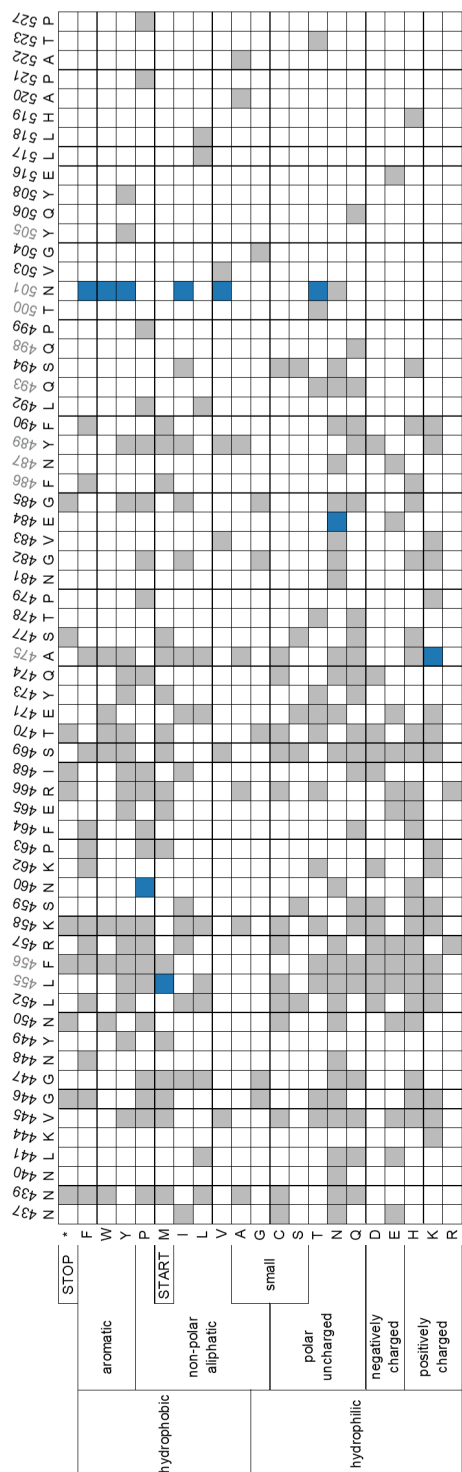
Figure C.11: **CC12.13 heatmap** using FDR <0.1 as cut-off. Blue - escape mutant hit, white – not a hit, grey – mutation not present. Position numbers in grey indicate ACE2 footprint. Related to **Figure 3.2**.
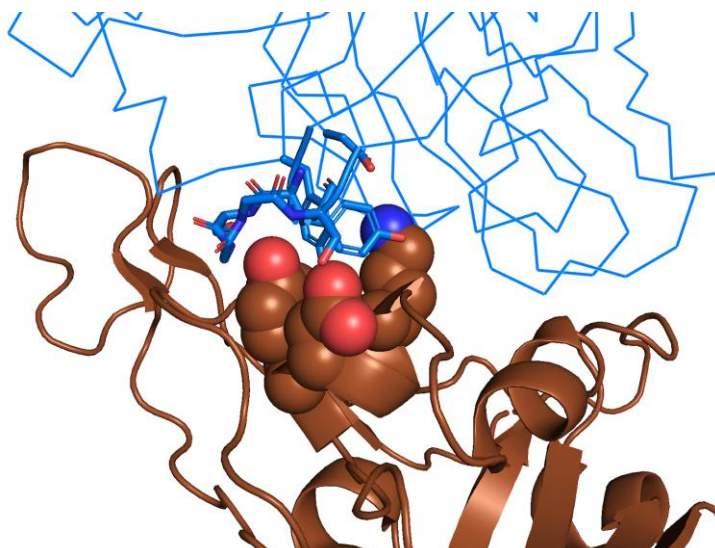
Figure C.12: **Structural recognition of S RBD (chocolate cartoon) by nAB CC12.1** (PDB ID 6XC2, blue ribbon). S positions K417, D420, Y421 are shown as spheres and the CDR H2 and key CC12.1 residues are shown as sticks. Related to **Figure 3.2**.

Figure C.13: **MLV-based SARS-CoV-2 pseudo-virus neutralization assays for SARS-CoV-2 RBD variants.** Data shown are replicates of the neutralization assays repeated on separate days (replicate 1 - day 1; replicate 2 - day 2). IC50 neutralization values shown are averages of two technical replicates. Replicate 1 data is also presented in Figure 1j of the main text. Related to **Figure 3.2**.
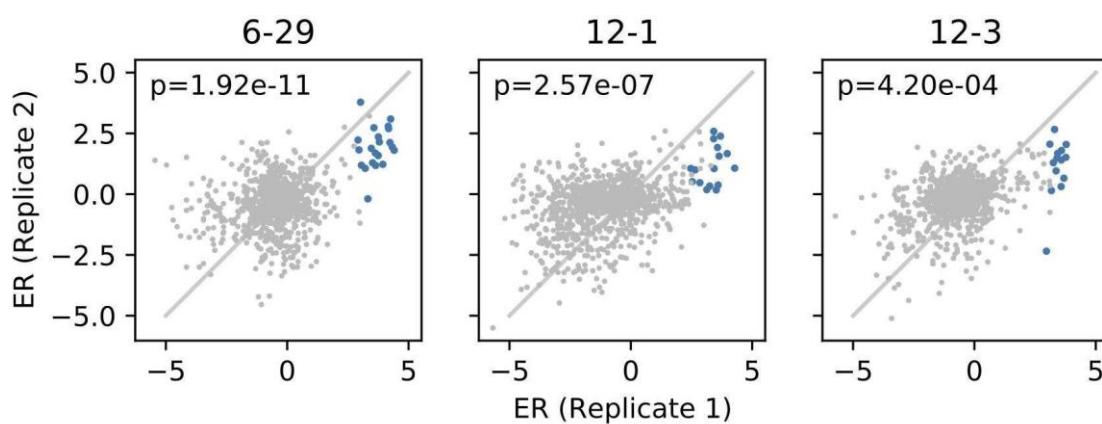
Figure C.14: **Comparison between biological replicates for S RBD positions 437-527 for CC6.29, CC12.1, and CC12.3.** Escape mutant hits identified in replicate 1 are shown as closed blue circles (a p $\leq$ 0.01 for an FDR <1). p-values are calculated using a one-sided Welch's t-test with the alternative hypothesis that the mean enrichment ratio from the replicate 2 hits are >the mean enrichment ratio from the replicate 2 non-hits. Related to **Figure 3.2**.
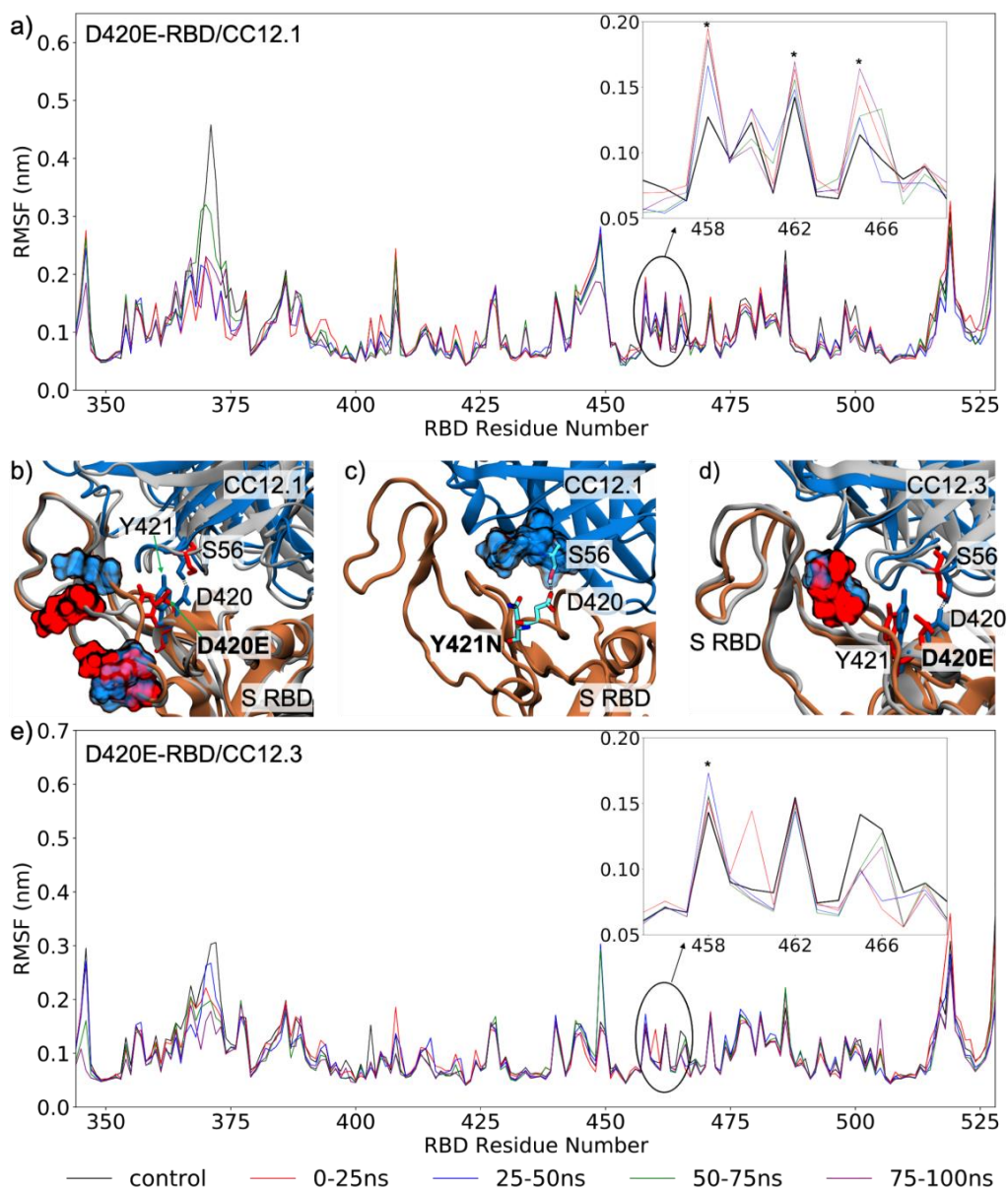
Figure C.15: (a) Comparison of the root mean square fluctuation (RMSF) profile of CC12.1 in complex with wildtype/control S RBD (averaged across the 100 ns production run) and S RBD with the D420E mutation (averaged across 25 ns intervals). (b) Structural mapping of highly fluctuating residues on S RBD with the D420E mutation, identified in panel (a), when complexed with CC12.1. Wildtype and mutant RBDs are shown in brown and gray and CC12.1 in blue and gray, respectively, while residues are colored blue for wildtype RBD and red for the mutant RBD; highly fluctuating residues are shown in surface representation. (c) MD snapshot showing the proposed mechanism of escape of S RBD from CC12.1 through mutation Y421N. (d) Structural mapping of highly fluctuating residues on S RBD with the D420E mutation, identified in panel (e), when complexed with CC12.3. (e) Comparison of the RMSF profile of CC12.3 in complex with wildtype S RBD and S RBD with the D420E mutation. Related to **Figure 3.4**
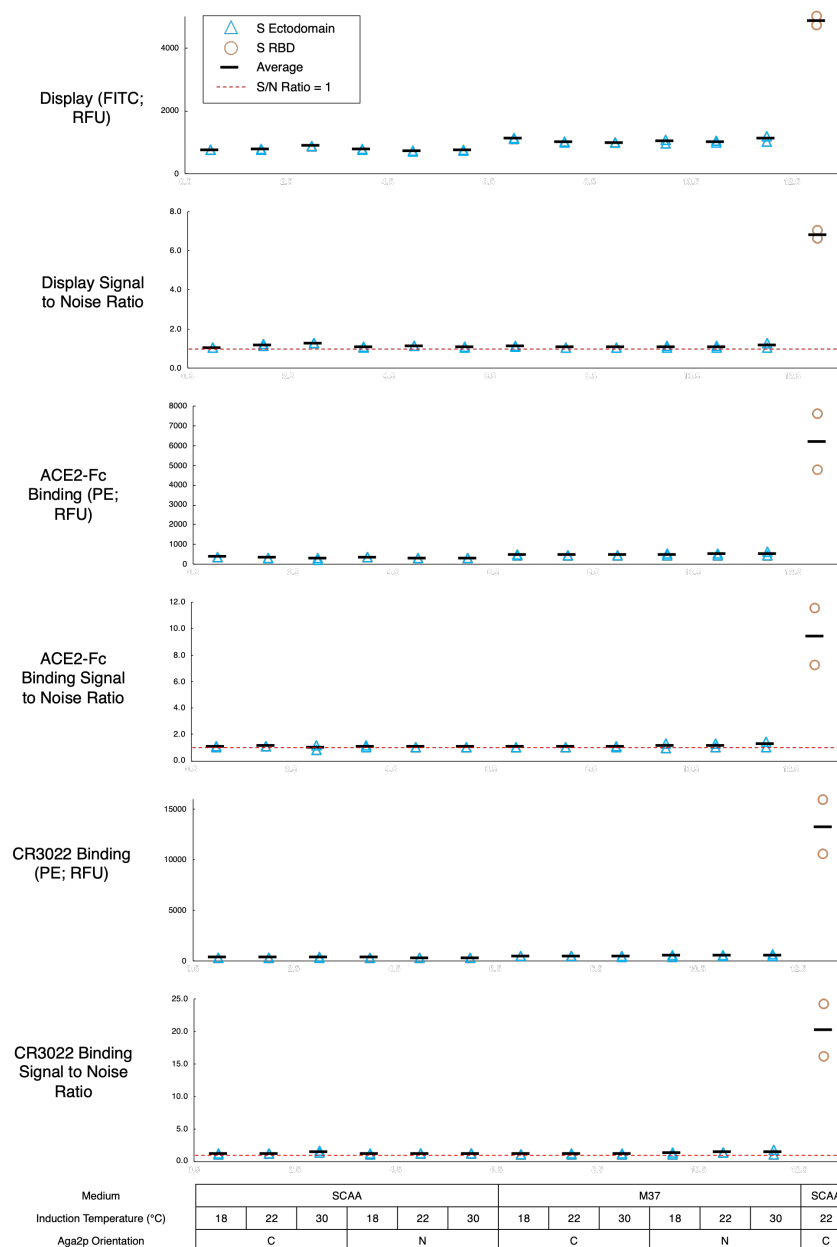
Figure C.16: **Yeast surface display for SARS-CoV-2 prefusion stabilized S ectodomain compared to S RBD.** FITC signal (RFU), FITC signal to noise ratio, and PE signal (RFU) and PE signal to noise ratio for both ACE2-Fc and CR3022 are shown for biological replicates of spike ectodomain with differing media, induction temperatures, and orientation of Spike relative to Aga2p (blue triangles). The FITC fluorescence derives from an anti-cmyc FITC antibody that recognizes a C-terminal cmyc epitope tag for displayed protein, while the PE signal is from biotinylated ACE2-Fc or CR3022 subsequently labeled with streptavidin-PE. The concentration of the secondary binding protein for the S ectodomain was 500nM ACE2-Fc and 500nM CR3022. FITC signal (RFU), FITC signal to noise ratio, and PE signal (RFU) and PE signal to noise ratio for both ACE2-Fc and CR3022 shown for S RBD at optimal conditions (orange circles). The concentration of secondary binding protein for the S RBD is at 1 nM, which is at saturation.
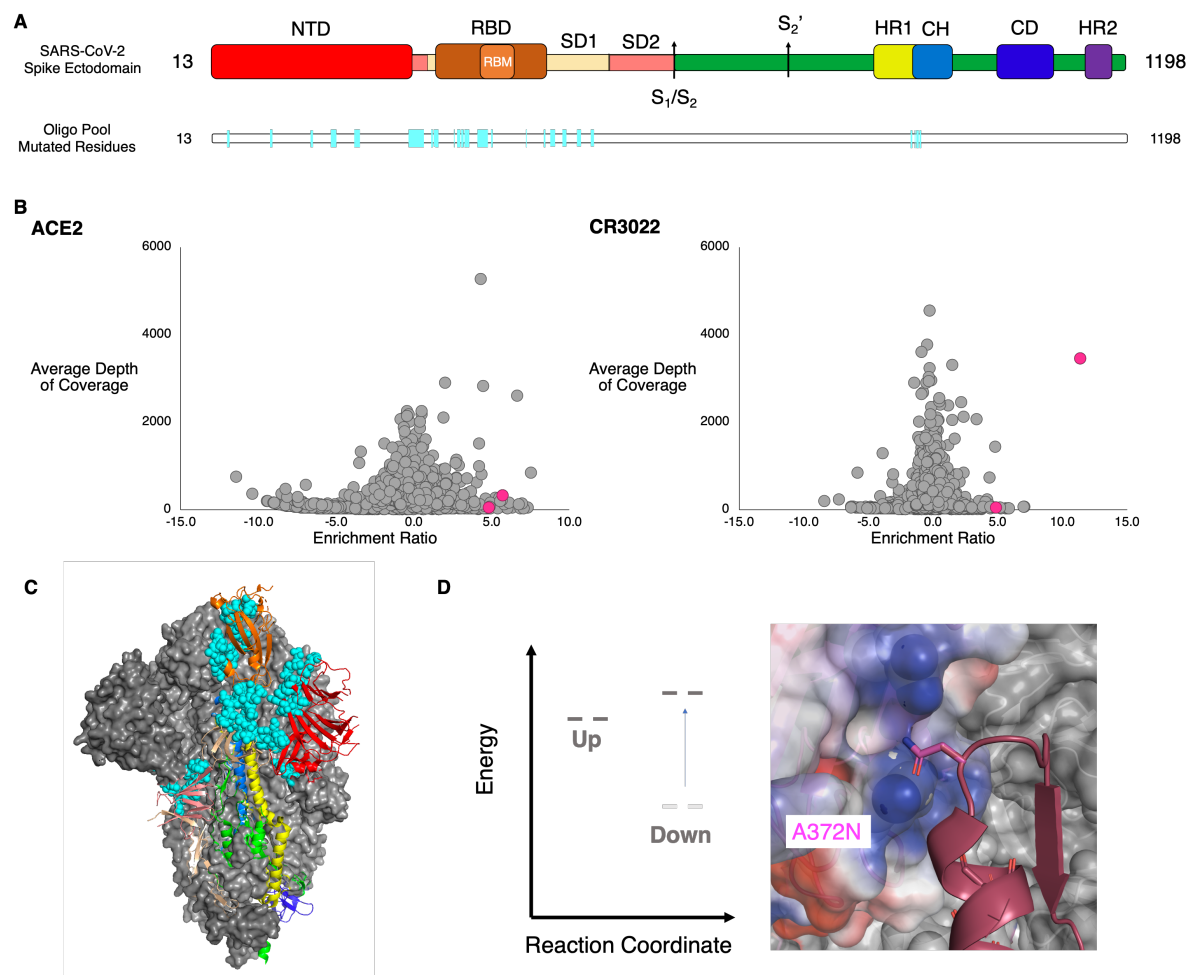
Figure C.17: **Determination and location of potential stabilizing hits.** (A) Spike ectodomain schematic with labeled and colored boundaries. Below schematic is the locations of the mutated residues in the oligo pool, cyan, as well as locations of the top identified hits shown in magenta. NTD: N-terminal domain, RBD: receptor-binding domain, RBM: receptor-binding motif, SD1: subdomain 1, SD2: subdomain 2, S1/S2: furin cleavage site, S2': S2' cleavage site, HR1: heptad repeat 1, CH: central helix, CD: connector domain, HR2: heptad repeat 2. (B) Average depth of coverage vs. Enrichment ratio for single mutants for (left) ACE2 and (right) CR3022. All mutations are shown in gray with the two hits (K113I & A372N) colored magenta. (C) Prefusion spike trimer shown with domains colored as they are in panel A. RBD is shown in the up conformation and along with SD1 and SD2 is shown on the same spike monomer. The NTD and S2 subunits are shown on a neighboring monomer. Oligo pool mutated residues are represented as cyan spheres. (D) Putative reaction coordinate and interaction of hit A372N in the spike ectodomain. A372N is hypothesized to destabilize the down protomer by steric repulsion with an adjacent 'down' protomer.
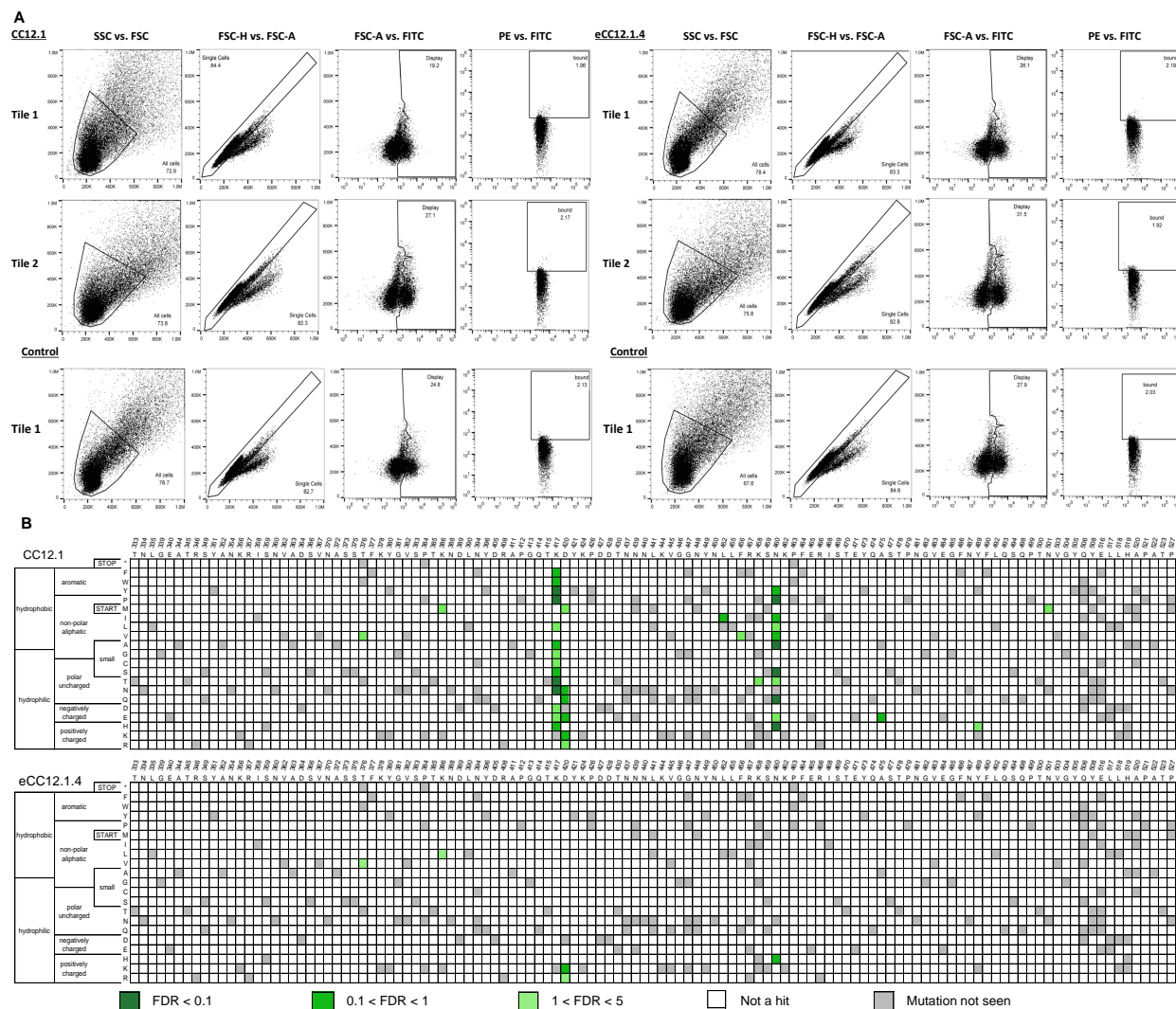
Figure C.18: **Identification of SARS-CoV-2 RBD escape mutants using yeast screening. Related to Figure 4.1. A.** Gating strategy and gates set for the escape mutant analysis for CC12.1 and eCC12.1.4. Control is the population of yeast cells without labeling by either biotinylated ACE2 or a given nAb. **B.** Heatmap showing predicted RBD (residues 333-527) escape mutants for CC12.1 and eCC12.1.4 in yellow to burnt orange with varying levels of confidence according to the key
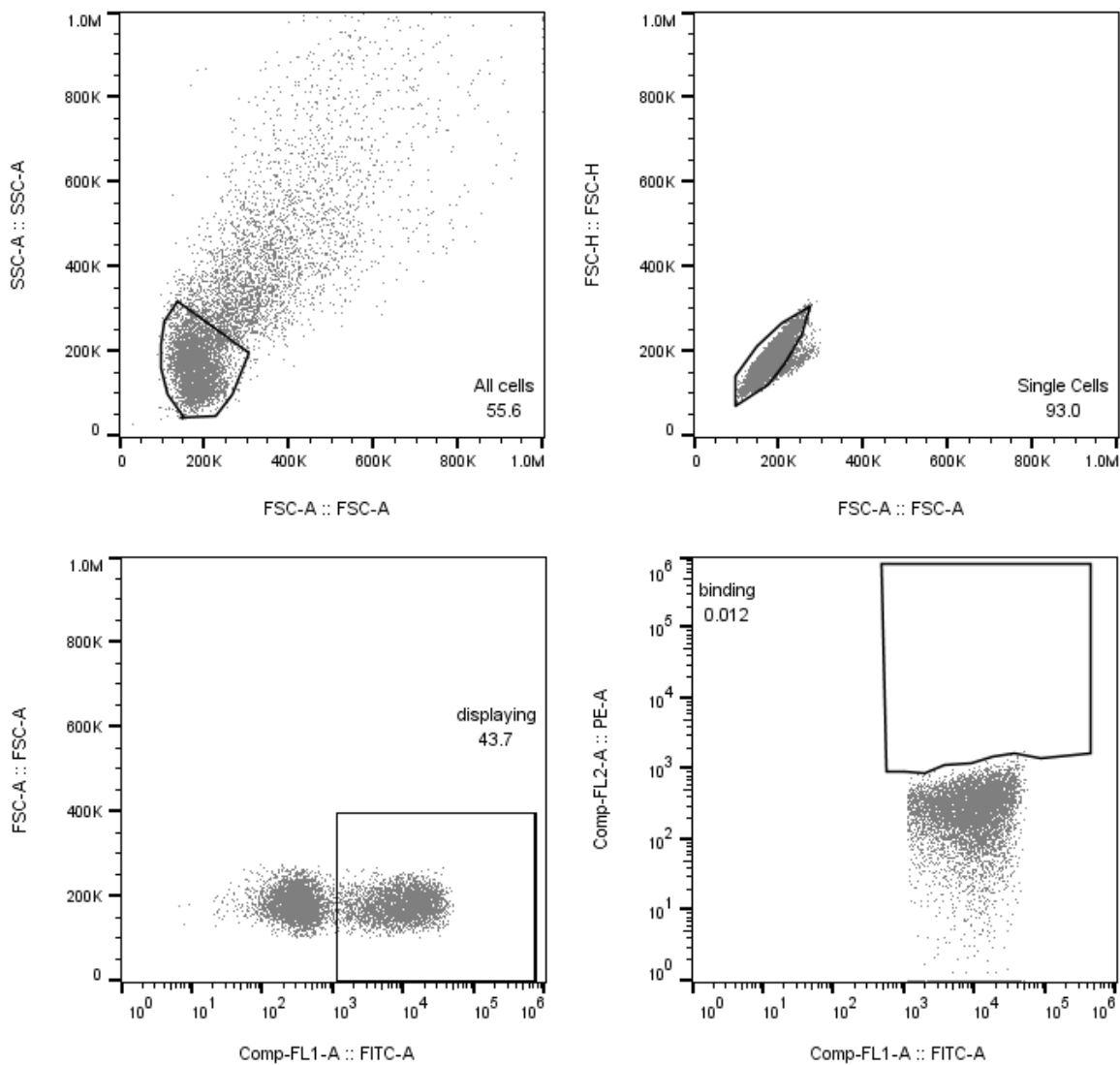
Figure C.19: **FACS sorting gates used to collect S RBD variants that maintain ACE2 binding.** Representative sorting gates used for all FACS screens. The three gates were SSC/FSC; FSC-H/FSC-A to discriminate single yeast cells; FSC-A/FITC$^+$ to select cells displaying the RBD on their surface; and SAPE$^+$/FITC$^+$ to identify mutants that allow ACE2 binding
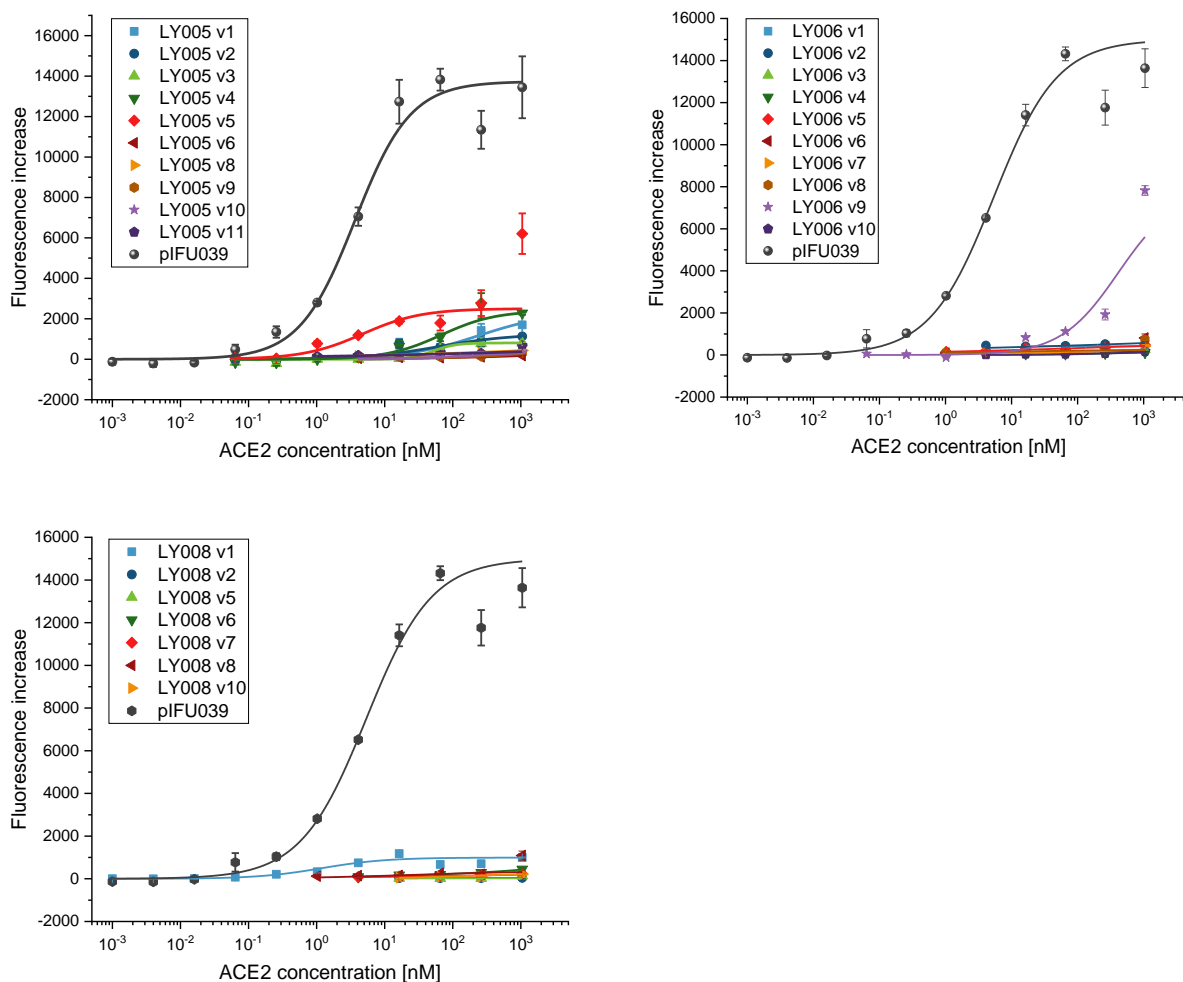
Figure C.20: **Isogenic titration curves.** Individual colonies were randomly selected from each library. For each variant, isogenic titrations were performed ranging from 1pM to 1μM ACE2. Titration curves are separated by library: LY005 (top left), LY006 (top right) and LY008 (bottom left). Wild type (pIFU039) titration curve is included as a control. Technical replicates were performed for each variant, error bars represent 2s.e.m.. A wild type titration was performed in duplicate each day, therefore, error bars for pIFU039 represent 4 s.e.m.
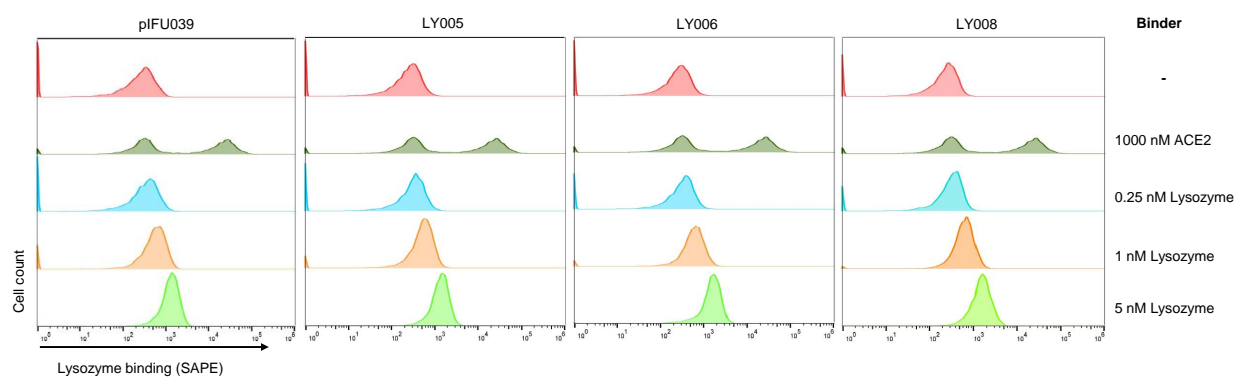
Figure C.21: **Polyspecificity assay.** The different libraries (LY005, LY006 and LY008) as well as WT (pIFU039) were labeled with different concentrations of lysozyme (0.25nM, 1nM and 5nM) for 30min at room temperature with shaking. As a control for specific binding we labeled the cells with 100nM ACE2 at the same conditions. We observed two peaks of SAPE fluorescence, one from the non-binding cells and the second from the cells displaying RBD that binds to ACE2. In the presence of lysozyme, a single peak of SAPE fluorescence is observed showing that the displayed RBD does not have non-specific binding. The fluorescence of the non-binding cells in the presence of lysozyme increases due to the binding of lysozyme to the yeast cell wall.

Appendix   D

Supplementary Tables

Table D.1: **Summary of Statistics for original Wuhan-Hu-1 S RBD N343Q & N-Term Spike ectodomain Libraries.** Library statistics for S RBD N343Q ("S RBD WT") and variants were determined by tabulating all variants with at least two reads in the yeast reference libraries. Library statistics for S ectodomain ("S Ecto") were determined from deep sequencing of the libraries harbored in E. coli. Related to Chapter 3 STAR Methods.

Tile Number	S RBD WT		S RBD E484K		S RBD N501Y		S Ecto				
	Tile 1	Tile 2	Tile 1	Tile 2	Tile 1	Tile 2	Tile 1	Tile 2	Tile 3	Tile 4	Tile 5
Positions	333-436	437-537	333-436	437-537	333-436	437-537	31-177	189-341	355-494	510-589	936-1020
Number of Designed Mutations	1120	1260	1120	1260	1120	1260	229	582	639	324	135
Transformants Obtained from Nicking Saturation Mutagenesis	1.10E+06	2.70E+07	3.00E+07	2.00E+06	1.00E+06	3.00E+07		2.00E+06		1.00E+06	
Transformants Obtained from Assembly of Spike Ectodomain	NA		NA		NA				4.00E+05		
Transformants Obtained from Homologous Recombination	2.50E+06	1.10E+06	4.00E+06	6.00E+06	3.00E+06	1.00E+06			1.75E+06		
Library Coverage Per Tile	97% (1085/1120)	94% (1189/1260)	90.5% (1014/1120)	87.1% (1097/1260)	90.5% (1014/1120)	92.9% (2185/2380)	88% (201/229)	90% (523/582)	90% (577/639)	100% (324/324)	99% (134/135)
Overall library coverage	96% (2274/2380)		88.7% (2111/2380)		91.8% (2185/2380)		92% (1759/1909)				

Table D.2: **List of escape mutants identified in this study.** Escape mutants were identified by a p value <0.01 for containing a FDR <1. Related to **Figure 3.1**.

nAb	Variant	Counts in competition selection	Counts in reference population	Enrichment Ratio	Minimum nucleotide distance	FDR	p value for FDR <1
CC.12-1	D405K	249	24	3.5	2	7.7E-07	1.5E-03
CC.12-1	Q414A	566	28	4.5	2	5.0E-13	8.3E-16
CC.12-1	K417A	634	19	5.2	2	5.0E-13	6.8E-25
CC.12-1	K417C	506	21	4.7	3	5.0E-13	1.9E-16
CC.12-1	K417D	358	7	5.8	2	5.0E-13	1.4E-17
CC.12-1	K417E	800	23	5.3	1	5.0E-13	1.0E-31
CC.12-1	K417F	446	12	5.4	3	5.0E-13	6.3E-19
CC.12-1	K417G	1655	54	5.1	2	5.0E-13	1.9E-59
CC.12-1	K417H	432	17	4.8	2	5.0E-13	8.5E-15
CC.12-1	K417I	155	12	3.8	1	1.3E-10	1.6E-03
CC.12-1	K417L	2303	63	5.3	2	5.0E-13	1.1E-90
CC.12-1	K417N	732	19	5.4	1	5.0E-13	1.1E-30
CC.12-1	K417P	1197	11	6.9	2	5.0E-13	3.0E-68
CC.12-1	K417S	1869	45	5.5	2	5.0E-13	1.1E-78
CC.12-1	K417T	1528	38	5.5	1	5.0E-13	1.2E-63
CC.12-1	K417V	787	33	4.7	2	5.0E-13	1.5E-24
CC.12-1	K417W	1427	42	5.2	2	5.0E-13	1.2E-54
CC.12-1	K417Y	523	15	5.3	2	5.0E-13	2.8E-21
CC.12-1	D420A	4199	84	5.8	1	5.0E-13	5.5E-189
CC.12-1	D420C	707	48	4.0	2	7.5E-13	1.1E-13

Table D.2 Continued

nAb	Variant	Counts in competition selection	Counts in reference population	Enrichment Ratio	Minimum nucleotide distance	FDR	p value for FDR <1
CC.12-1	D420E	3842	113	5.2	1	5.0E-13	2.7E-144
CC.12-1	D420H	372	34	3.6	1	1.1E-07	3.9E-05
CC.12-1	D420K	584	10	6.0	2	5.0E-13	4.0E-29
CC.12-1	D420L	1885	54	5.3	2	5.0E-13	9.8E-73
CC.12-1	D420M	756	26	5.0	3	5.0E-13	3.9E-27
CC.12-1	D420N	914	24	5.4	1	5.0E-13	1.0E-37
CC.12-1	D420Q	678	24	5.0	2	5.0E-13	4.8E-24
CC.12-1	D420R	4431	110	5.5	2	5.0E-13	1.8E-181
CC.12-1	D420S	1979	49	5.5	2	5.0E-13	4.0E-82
CC.12-1	D420T	733	33	4.6	2	5.0E-13	1.0E-21
CC.12-1	D420W	372	41	3.3	3	5.8E-05	1.5E-03
CC.12-1	D420Y	299	24	3.8	1	6.4E-10	2.4E-05
CC.12-1	Y421A	888	109	3.2	2	1.1E-03	9.5E-05
CC.12-1	Y421C	1952	126	4.1	1	5.0E-13	1.9E-37
CC.12-1	Y421H	1140	27	5.5	1	5.0E-13	5.6E-49
CC.12-1	Y421L	5137	159	5.2	2	5.0E-13	1.1E-186
CC.12-1	Y421N	1262	56	4.6	1	5.0E-13	1.1E-36
CC.12-1	Y421W	3129	156	4.5	2	5.0E-13	1.9E-79
CC.12-1	Q498H	4086	296	3.6	1	1.7E-02	6.9E-30
CC.12-1	Q498Y	1467	147	3.1	2	3.7E-01	2.0E-04
CC.12-1	N501F	7508	617	3.4	2	7.6E-02	3.0E-37
CC.12-1	N501M	9286	519	4.0	2	2.0E-04	4.0E-111
CC.12-1	N501T	10942	890	3.4	1	6.8E-02	3.1E-55

Table D.2 Continued

nAb	Variant	Counts in competition selection	Counts in reference population	Enrichment Ratio	Minimum nucleotide distance	FDR	p value for FDR <1
CC.12-1	N501V	12572	885	3.6	2	1.1E-02	1.7E-94
CC.12-1	N501W	4503	305	3.7	3	6.3E-03	2.9E-38
CC.12-1	N501Y	7767	639	3.4	1	7.6E-02	2.5E-38
CC.12-1	Y505W	12149	894	3.6	2	2.1E-02	9.4E-82
CC.12-1	Y508H	1478	113	3.5	1	3.3E-02	1.8E-10
CC.12-1	P527I	467	43	3.3	2	2.1E-01	7.4E-03
CC.12-1	P527M	155	7	4.3	2	1.3E-06	7.5E-04
CC.12-3	Q414A	627	28	4.4	2	5.0E-13	1.1E-13
CC.12-3	Q414G	807	44	4.1	2	5.0E-13	9.8E-14
CC.12-3	T415G	1424	82	4.0	2	1.4E-12	2.3E-21
CC.12-3	K417A	514	19	4.6	2	5.0E-13	1.7E-13
CC.12-3	K417C	496	21	4.4	3	5.0E-13	1.1E-11
CC.12-3	K417D	210	7	4.8	2	5.0E-13	1.0E-06
CC.12-3	K417E	557	23	4.5	1	5.0E-13	3.1E-13
CC.12-3	K417F	344	12	4.7	3	5.0E-13	6.8E-10
CC.12-3	K417G	1439	54	4.6	2	5.0E-13	1.3E-34
CC.12-3	K417H	352	17	4.2	2	5.0E-13	1.0E-07
CC.12-3	K417I	502	12	5.3	1	5.0E-13	2.3E-17
CC.12-3	K417L	2239	63	5.0	2	5.0E-13	1.2E-65
CC.12-3	K417M	983	23	5.3	1	5.0E-13	5.4E-33
CC.12-3	K417N	467	19	4.5	1	5.0E-13	1.8E-11
CC.12-3	K417P	922	11	6.3	2	5.0E-13	3.5E-40
CC.12-3	K417Q	680	28	4.5	1	5.0E-13	7.1E-16

Table D.2 Continued

nAb	Variant	Counts in competition selection	Counts in reference population	Enrichment Ratio	Minimum nucleotide distance	FDR	p value for FDR <1
CC.12-3	K417R	1486	84	4.0	1	8.7E-13	8.3E-23
CC.12-3	K417S	1475	45	4.9	2	5.0E-13	1.5E-41
CC.12-3	K417T	1229	38	4.9	1	5.0E-13	1.5E-34
CC.12-3	K417V	1080	33	4.9	2	5.0E-13	8.3E-31
CC.12-3	K417W	1127	42	4.6	2	5.0E-13	1.3E-27
CC.12-3	K417Y	344	15	4.4	2	5.0E-13	2.6E-08
CC.12-3	D420A	2677	84	4.9	1	5.0E-13	1.5E-72
CC.12-3	D420C	619	48	3.6	2	3.0E-07	2.5E-06
CC.12-3	D420E	2875	113	4.5	1	5.0E-13	1.0E-64
CC.12-3	D420K	317	10	4.9	2	5.0E-13	7.6E-10
CC.12-3	D420L	1313	54	4.5	2	5.0E-13	3.1E-29
CC.12-3	D420M	550	26	4.3	3	5.0E-13	1.6E-11
CC.12-3	D420N	595	24	4.5	1	5.0E-13	2.5E-14
CC.12-3	D420Q	410	24	4.0	2	2.6E-12	4.6E-07
CC.12-3	D420R	2769	110	4.5	2	5.0E-13	8.6E-62
CC.12-3	D420S	1061	49	4.3	2	5.0E-13	2.4E-21
CC.12-3	D420T	530	33	3.9	2	4.0E-11	5.1E-08
CC.12-3	D420Y	273	24	3.4	1	2.0E-05	7.1E-03
CC.12-3	Y421A	1030	109	3.1	2	3.1E-03	4.8E-04
CC.12-3	Y421F	729	80	3.1	1	7.2E-03	6.7E-03
CC.12-3	Y421H	881	27	4.9	1	5.0E-13	2.3E-25
CC.12-3	Y421L	3735	159	4.4	2	5.0E-13	2.6E-77
CC.12-3	Y421N	1039	56	4.1	1	5.0E-13	1.7E-17

Table D.2 Continued

nAb	Variant	Counts in competition selection	Counts in reference population	Enrichment Ratio	Minimum nucleotide distance	FDR	p value for FDR <1
CC.12-3	Y421W	3746	156	4.5	2	5.0E-13	2.9E-79
CC.12-3	N460V	315	28	3.4	2	1.1E-01	8.5E-03
CC.12-3	Q498H	3844	296	3.6	1	2.1E-02	4.3E-28
CC.12-3	Q498W	4847	466	3.3	2	2.1E-01	4.7E-17
CC.12-3	Q498Y	1445	147	3.2	2	3.1E-01	4.8E-05
CC.12-3	N501F	9266	617	3.8	2	2.3E-03	7.4E-91
CC.12-3	N501W	4284	305	3.7	3	6.9E-03	1.9E-37
CC.12-3	N501Y	9508	639	3.8	1	2.7E-03	1.9E-91
CC.12-3	Y505W	11672	894	3.6	2	1.9E-02	1.3E-82
CC.12-13	A475K	35	11	3.3	2	3.1E-06	3.7E-02
CC.12-13	N501W	2074	745	3.1	3	1.2E-04	1.3E-35
CC.12-13	N460P	21	8	3.0	2	5.1E-04	1.8E-01
CC.12-13	E484N	21	8	3.0	2	5.1E-04	1.8E-01
CC.12-13	N501F	3378	1311	3.0	2	7.8E-04	3.5E-43
CC.12-13	N501V	4041	1593	3.0	2	1.1E-03	3.5E-48
CC.12-13	N501Y	3028	1339	2.8	1	1.2E-02	2.4E-21
CC.12-13	L455M	19	10	2.6	1	2.1E-01	4.5E-01
CC.12-13	N501T	3053	1677	2.5	1	3.6E-01	2.1E-03
CC.12-13	N501I	2695	1531	2.4	1	5.5E-01	4.9E-02
CC.6-29	A475R	174	6	4.4	2	1.1E-07	5.6E-04
CC.6-29	S477P	472	18	4.3	1	1.8E-06	7.2E-08
CC.6-29	T478L	1562	85	3.7	2	3.4E-03	2.1E-13
CC.6-29	T478Q	239	12	3.9	2	8.4E-04	1.6E-03

Table D.3: **Mutations identified in literature.** [30, 31, 84, 85, 86]. Related to **Figure 3.4**.

Mutation	Antibody	Germline	Source (PMID)
T345A	2H04	IGHV1-55	33535027
T345N	2H04	IGHV1-55	33535027
T345S	2H04	IGHV1-55	33535027
R346G	SARS2-01		33535027
R346G	2H04	IGHV1-55	33535027
R346K	C135	IGHV3-30	32743579
R346M	C135	IGHV3-30	32743579
R346S	C135	IGHV3-30	32743579
A352D	SARS2-01		33535027
Y369C	COV2-2082	IGHV3-20	32935107
N370K	COV2-2082	IGHV3-20	32935107
N370S	COV2-2082	IGHV3-20	32935107
A372S	COV2-2082	IGHV3-20	32935107
A372T	COV2-2082	IGHV3-20	32935107
A372V	COV2-2082	IGHV3-20	32935107
T376I	COV2-2082	IGHV3-20	32935107
T376I	COV2-2094	IGHV3-20	32935107
K378E	SARS2-31		33535027
K378N	COV2-2677	IGHV4-39	32935107
K378N	COV2-2082	IGHV3-20	32935107
K378N	COV2-2094	IGHV3-20	32935107
K378Q	COV2-2677	IGHV4-39	33535027
K378R	COV2-2677	IGHV4-39	32935107

Table D.3 Continued

Mutation	Antibody	Germline	Source (PMID)
K378R	COV2-2082	IGHV3-20	32935107
K378R	COV2-2094	IGHV3-20	32935107
P384L	COV2-2677	IGHV4-39	32935107
P384S	COV2-2677	IGHV4-39	32935107
R408I	COV2-2082	IGHV3-20	32935107
R408I	COV2-2094	IGHV3-20	32935107
R408K	COV2-2082	IGHV3-20	32935107
R408K	COV2-2094	IGHV3-20	32935107
R408K	SARS2-31		33535027
R408T	COV2-2082	IGHV3-20	32935107
R408T	COV2-2094	IGHV3-20	32935107
A411S	COV2-2082	IGHV3-20	32935107
K417E	REGN10933	IGHV3-48	32540904
K417N	COV2-2082	IGHV3-20	32935107
K417N	COV2-2094	IGHV3-20	32935107
K417R	COV2-2082	IGHV3-20	32935107
K417R	COV2-2094	IGHV3-20	32935107
A435S	COV2-2094	IGHV3-20	32935107
D614G+A435S	H014		32730807
N439K	H00S022		32730807
N440K	C135	IGHV3-30	32743579
L441R	2H04	IGHV1-55	33535027
K444E	2H04	IGHV1-55	33535027
K444E	SARS2-38		33535027

Table D.3 Continued

Mutation	Antibody	Germline	Source (PMID)
K444E	SARS2-22		33535027
K444N	SARS2-38		33535027
K444R	SARS2-22		33535027
K444Q	REGN10934	IGHV3-15	32540904
K444Q	REGN10987	IGHV3-30	32540904
K444R	SARS2-22		33535027
V445A	REGN10987	IGHV3-30	32540904
V445A	COV2-2499	IGHV4-39	32935107
V445F	COV2-2499	IGHV4-39	32935107
V445G	SARS2-22		33535027
V445I	COV2-2499	IGHV4-39	32935107
G446A	COV2-2096	IGHV1-8	32935107
G446A	COV2-2499	IGHV4-39	32935107
G446D	SARS2-02		33535027
G446D	SARS2-32		33535027
G446D	SARS2-38		33535027
G446D	SARS2-22		33535027
G446S	COV2-2096	IGHV1-8	32935107
G446S	COV2-2499	IGHV4-39	32935107
G446V	COV2-2096	IGHV1-8	32935107
G446V	COV2-2499	IGHV4-39	32935107
G446V	SARS2-02		33535027
N450D	SARS2-07		33535027
N450K	SARS2-32		33535027

Table D.3 Continued

Mutation	Antibody	Germline	Source (PMID)
N450Y	SARS2-32		33535027
L452M	COV2-2096	IGHV1-8	32935107
L452R	COV2-2096	IGHV1-8	32935107
L452R	X593		32730807
L452R	P2B-2F6	IGHV4-38	32730807
L452R	SARS2-01		33535027
L452R	SARS2-32		33535027
Y453F	REGN10933	IGHV3-48	32540904
L455F	REGN10933	IGHV3-48	32540904
K458Q	SARS2-66		33535027
I472V	COV2-2479	IGHV1-69	32935107
D614G+I472V	X593		32730807
Q474P	SARS2-34		33535027
A475V	157		32730807
A475V	247		32730807
A475V	CB6	IGHV3-66	32730807
A475V	P2C-1F11	IGHV3-66	32730807
A475V	B38	IGHV3-53	32730807
A475V	CA1	IGHV1-18	32730807
A475V	COV2-2165	IGHV3-66	32935107
A475V	COV2-2832	IGHV3-66	32935107
G476D	SARS2-21		33535027
G476D	SARS2-34		33535027
G476D	SARS2-71		33535027

Table D.3 Continued

Mutation	Antibody	Germline	Source (PMID)
G476S	SARS2-21		33535027
S477G	SARS2-16		33535027
S477G	SARS2-07		33535027
S477G	SARS2-19		33535027
S477G	SARS2-34		33535027
S477G	SARS2-58		33535027
S477G	SARS2-71		33535027
S477I	SARS2-58		33535027
S477N	SARS2-16		33535027
S477N	SARS2-07		33535027
S477N	SARS2-19		33535027
S477N	SARS2-34		33535027
S477N	SARS2-58		33535027
P477L	SARS2-07		33535027
S477R	SARS2-07		33535027
S477R	SARS2-16		33535027
S477R	SARS2-23		33535027
S477R	SARS2-34		33535027
T478I	SARS2-19		33535027
T478I	SARS2-21		33535027
T478I	SASRS2-71		33535027
T478P	SASRS2-71		33535027
P479L	SARS2-34		33535027
P479L	SARS2-71		33535027

Table D.3 Continued

Mutation	Antibody	Germline	Source (PMID)
P479S	SARS2-21		33535027
V483A	X593		32730807
V483A	P2B-2F6	IGHV4-38	32730807
V483F	SARS2-23		33535027
V483G	SARS2-23		33535027
E484A	2B04	IGHV2-9	33535027
E484A	COV2-2832	IGHV3-66	32935107
E484A	COV2-2479	IGHV1-69	32935107
E484A	COV2-2050	IGHV1-2	32935107
E484A	1B07	IGHV2-9	33535027
E484A	COV2-2096	IGHV1-8	32935107
E484D	COV2-2832	IGHV3-66	32935107
E484D	COV2-2479	IGHV1-69	32935107
E484D	COV2-2050	IGHV1-2	32935107
E484D	COV2-2096	IGHV1-8	32935107
E484D	1B07	IGHV2-9	33535027
E484D	SARS2-23		33535027
E484D	SARS2-66		33535027
E484G	1B07	IGHV2-9	33535027
E484K	COV2-2832	IGHV3-66	32935107
E484K	COV2-2479	IGHV1-69	32935107
E484K	COV2-2050	IGHV1-2	32935107
E484K	COV2-2096	IGHV1-8	32935107
E484K	REGN10989	IGHV1-2	32540904

Table D.3 Continued

Mutation	Antibody	Germline	Source (PMID)
E484K	REGN10933	IGHV3-48	32540904
E484K	REGN10934	IGHV3-15	32540904
E484K	REGN10989/10934	IGHV1-2/IGHV3-15	32540904
E484K	2B04	IGHV2-9	33535027
E484K	1B07	IGHV2-9	33535027
E484K	SARS2-02		33535027
E484K	SARS2-32		33535027
E484K	SARS2-58		33535027
E484K	C121	IGHV1-2	32743579
E484K	C144	IGHV3-53	32743579
E484Q	COV2-2832	IGHV3-66	32935107
E484Q	COV2-2479	IGHV1-69	32935107
E484Q	COV2-2050	IGHV1-2	32935107
E484Q	COV2-2096	IGHV1-8	32935107
G485D	REGN10989	IGHV1-2	32540904
F486S	2B04	IGHV2-9	33535027
F486S	SARS2-21		33535027
F486V	REGN10989	IGHV1-2	32540904
F486V	REGN10933	IGHV3-48	32540904
F486V	SARS2-58		33535027
F486V	SARS2-71		33535027
F486L	COV2-2832	IGHV3-66	32935107
F486L	SARS2-21		33535027
F486Y	1B07	IGHV2-9	33535027

Table D.3 Continued

Mutation	Antibody	Germline	Source (PMID)
F490P	REGN10989	IGHV1-2	32540904
F490P	REGN10934	IGHV3-15	32540904
F490P	REGN10989/10934	IGHV1-2/IGHV3-15	32540904
F490L	X593		32730807
F490L	261-262		32730807
F490L	H4	IGHV1-2	32730807
F490L	P2B-2F6	IGHV4-38	32730807
F490L	COV2-2479	IGHV1-69	32935107
F490L	COV2-2050	IGHV1-2	32935107
F490L	COV2-2096	IGHV1-8	32935107
F490L	SARS2-66		33535027
F490L	C121	IGHV1-2	32743579
F490S	COV2-2479	IGHV1-69	32935107
F490S	COV2-2050	IGHV1-2	32935107
F490S	COV2-2096	IGHV1-8	32935107
F490S	SARS2-32		33535027
Q493K	REGN10989	IGHV1-2	32540904
Q493K	REGN10933	IGHV3-48	32540904
Q493K	REGN10989/10934	IGHV1-2/IGHV3-15	32540904
Q493K	C144	IGHV3-53	32743579
Q493R	C144	IGHV3-53	32743579
Q493K	C121	IGHV1-2	32743579
S494L	COV2-2096	IGHV1-8	32935107
S494P	COV2-2096	IGHV1-8	32935107

Table D.3 Continued

Mutation	Antibody	Germline	Source (PMID)
S494P	SARS2-01		33535027
P499H	COV2-2499	IGHV4-39	32935107
P499R	COV2-2499	IGHV4-39	32935107
P499S	COV2-2499	IGHV4-39	32935107
G504D	SARS2-31		33535027
Y508H	H014		32730807

Table D.4: **Table of all Golden Gate reactions performed**

Experiment	Rep.	Destination Vector	Insert(s)	Input DNA (fmol)	Size (μl)	Colonies	% GFP Negative	pUC19 Colonies /μg DNA
Transformants / Cassette Number (Fig. 5.2A)	1	pND003	PYR1 Cass1+2+3+4	40	25	8.4×10^5	99.9	3.7×10^9 (Rep. 1) 4.5×10^9 (Rep. 2)
	2		PYR1 Cass1+2+3+4			8.4×10^5	99.9	
	1		PYR1 Cass1+2+3, PYR1 Cass4			4.9×10^5	99.8	
	2		PYR1 Cass1+2+3, PYR1 Cass4			7.2×10^5	99.9	
	1		PYR1 Cass1+2, PYR1 Cass3, PYR1 Cass4			3.4×10^5	99.4	
	2		PYR1 Cass1+2, PYR1 Cass3, PYR1 Cass4			6.0×10^5	99.9	
	1		PYR1 Cass1, 2, 3, 4			3.5×10^5	99.7	
	2		PYR1 Cass1, 2, 3, 4			3.9×10^5	99.7	
GG RBD Function Test (Fig.5.1D,5.1E)	1	pIFU037	RBD WT eBlock (1+2+3)	40	20	3×10^3	80.0	NA
	2					1×10^4	91.6	
	3					1.1×10^6	96.0	
	4					5.3×10^5	98.0	
Transformants / Reaction Size (Fig. 5.2B)	1	pND003	PYR1 Cass1+2+3+4	40	25	3.2×10^5	96.9	8.3×10^9 (Rep. 1) 4.9×10^9 (Rep. 2) 7.1×10^9 (Rep. 3)
	2	pND003	PYR1 Cass1+2+3+4	40	25	9×10^5	98.9	
	3	pND003	PYR1 Cass1+2+3+4	40	25	9×10^5	98.9	
	1	pND003	PYR1 Cass1+2+3+4	160	100	2×10^6	99.5	
	2	pND003	PYR1 Cass1+2+3+4	160	100	2.4×10^6	99.6	
	3	pND003	PYR1 Cass1+2+3+4	160	100	2.7×10^6	99.6	
	1	pND003	PYR1 Cass1+2+3+4	320	200	8.2×10^6	99.9	
	2	pND003	PYR1 Cass1+2+3+4	320	200	2.3×10^7	99.9	
	3	pND003	PYR1 Cass1+2+3+4	320	200	1.9×10^7	99.9	
Transformants / Input DNA (Fig. 5.2C)	1	pND003	PYR1 Cass1+2+3+4	80	25	7x105	98.6	NA
	2	pND003	PYR1 Cass1+2+3+4	80	25	2.9×10^6	99.7	
	3	pND003	PYR1 Cass1+2+3+4	80	25	4×10^6	99.8	
	1	pND003	PYR1 Cass1+2+3+4	160	25	5.8×10^6	99.7	
	2	pND003	PYR1 Cass1+2+3+4	160	25	2.7×10^7	99.8	

	3	pND003	PYR1 Cass1+2+3+4	160	25	6.2×10^6	99.8	
RBD Library (Fig. 5.3A)	1	pIFU037	RBD ultramer 1+2+3	200	25	5.6×10^6	99.9	NA
	1					2×10^6	99.9	
	1		RBD oligo pool 1+2+3	400	50	1.6×10^7	98.0	
	2					2.3×10^7	92.0	
eBlock, Ultramer, oligo pool comparison (Fig. 5.3D)	1	pIFU037	WT RBD eBlock 1+2+3	40	25	1.5×10^7	99.9	6.2×10^9 (Rep. 1)
	2					2.5×10^6	99.9	
	3					3.4×10^6	99.9	
	1		WT RBD Ultramer 1+2+3			3.4×10^6	99.9	5×10^9 (Rep. 2)
	2					1.5×10^6	99.9	
	3					1.4×10^6	99.9	
	1		RBD oligo pool 1+2+3			3.0×10^6	99.9	2.5×10^9 (Rep. 3)
	2					1.1×10^6	99.9	
	3					2.2×10^6	99.9	

Table D.5: **Library statistics for S RBD sequenced libraries**

	Library 1	**Library 2**	**Library 3**
DNA type	Ultramer	Ultramer	Oligo pool
Number of designed mutations	7776	20736	17725
Number of transformants obtained	5.6×10^6	2×10^6	2×10^7
Cassette incorporation efficiency	99.9%	99.9%	99.9%
Library coverage	100% (7776/7776)	96.7% (20043/20736)	99.9% (17701/17725)
% desired reads (designed variants - right mutation and number of expected mutations)	86.8% (824718/949555)	83.9% (594753/708766)	80.9% (2607572/3224158)
% chimera reads (variants with the correct mutations but with more than 3 mutations in cassette 3)	0% (0/949555)	0% (0/708766)	8.9% (289648/3224158)
% WT sequence reads	0.22% (1798/824718)	0.06% (347/594753)	0.02% (483/2607572)

Table D.6: **S RBD Single saturation mutagenesis library coverage statistics.** Related to Figure 4.1.

	S RBD N343Q (Wuhan-1)	
Tile Number	Tile 1	Tile 2
Positions	333-436	437-537
Number of Designed Mutations	1120	1260
Transformants Obtained from Nicking Saturation Mutagenesis	5.00E+05	1.50E+06
Transformants Obtained from Homologous Recombination	8.00E+05	1.00E+06
Library Coverage Per Tile	97% (1089/1120)	92% (1161/1260)
Overall Library Coverage	94.5% (2250/2380)	

Table D.7: **Transformation efficiencies for the libraries generated in the study of RBD sequence varaibility**

	LY005	LY006	LY008	LY008 rep	LY009
Number of designed mutations	7776	20736	1611		2268
E. Coli Number of transformants	1.60E+06	1.30E+07	1.00E+09		1.00E+09
Cassette incorporation efficiency	99.99%	99.99%	99.99%		99.00%
Yeast Number of transformants	>1E+06	>1E+06	>1E+06	>1E+06	>1E+06
Library coverage	100% (7776/7776)	99.6% (20660/20736)	99.6% (1605/1611)	99.1% (1597/1611)	100% (2268/2268)