The Dissertation Committee for Jule Goike Certifies that this is the approved version of the following Dissertation:

METHODS IN ANTIBODY DISCOVERY

Committee:

George Georgiou, Supervisor

Jimmy D. Gollihar, Co-Supervisor

Andrew Ellington

Jennifer Maynard

Edward Marcotte

METHODS IN ANTIBODY DISCOVERY

by

Jule Goike

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2022

Dedication

To my mother

Thank you for nurturing my inquisitiveness and teaching me perseverance in the face of obstacles.

Acknowledgements

I would like to thank and acknowledge the people who have been instrumental along my academic journey. At UT, I am thankful to my supervisors Dr Gollihar and Dr Georgiou for giving me the opportunity to join their labs and carve my own path. I thank my committee members Dr Marcotte, Dr Maynard and Dr Ellington for insightful conversations and guidance throughout the years. I thoroughly enjoyed my rotation with Dr Ross Thyer and thank him for his mentorship, as well as Dr Andrew Horton and Dr Daniel Boutz for their advice and help throughout my Ph.D. I want to thank my co-workers Jonny Riggs, Ivan Gonzales, Dr Ebru Cayir, and Dr Donjeta Gjuka for their comradery and support while making the day-to-day lab life a lot more fun.

I want to express my appreciation and gratitude to Dr Lloyd-Evans, of Cardiff University, whose lab I had the privilege to be a part of as a placement student and who first ignited my passion for research through his excellent mentorship. Without my time in his lab, I would not have pursued a graduate degree. I would like to thank Herr Schnurr for his support and mentorship throughout high-school.

My special thanks goes to my friends throughout the years whose support was invaluable. I would like to acknowledge Luise Kirchhof, Roisin O'Brien, Foteini Bartzoka, Dr Martina Audagnotta, and Marija Petrovic for their friendship which has lifted me through difficult times. May we be the change we want to see.

I need to express my deep gratitude to my family for their unwavering support of my life decisions, even when they take me to opposite ends of the globe. Thank you so much for believing in me and my capabilities. Without your support and trust I could not have flourished. I thank my brother for always letting me rope him into long-winded

discussion on any and all topics, and for shouldering a lot of my responsibilities in my absence. I thank my mother for teaching me about the world and always trying to find answers to my many questions. Without your investment in me, I could not have developed into the person I am today, and for that I am thankful.

Last but not least, I want to thank my husband, Chris Falco, for his love and support throughout this journey. I love our conversations and deeply value your perspective and insights. Meeting you was a truly unexpected blessing of my Ph.D and what made it all worth it in the end.

Abstract

Methods in Antibody Discovery

Jule Goike, Ph.D

The University of Texas at Austin, 2022

Supervisors: George Georgiou, Jimmy D. Gollihar

Antibodies are an essential part of adaptive immunity and can be raised against almost any molecule. The ability to bind a wide variety of molecules with exquisite specificity, causes antibodies to be highly desirable tools for research and in medicine. Beginning with the discovery of antibodies in 1890 by Behring and Kitasato, the scientific community was aware of the enormous therapeutic potential of molecules that can be induced to bind any disease-causing factor. However, the realization of this potential was stifled by the lack of techniques to isolate and produce antibodies of defined specificity. Since then, versatile antibody discovery methods have been developed, however, initial antibody isolation and characterization remain a bottleneck in the antibody discovery process. The work presented here improves upon existing methods of antibody discovery. Development of a novel yeast surface display strain reduces selection times by half and expedites downstream development by use of a native antibody format that translates effortlessly into full-length antibodies. The combination of antibody yeast surface display and mass-spectrometry based protein sequencing of serum antibodies successfully isolates SARS-CoV-2 binders from infected donors. Additionally, examination of the role of antibody light chains leads to insights into repertoire diversity

and sparks an innovative approach to antibody heavy and light chain pairing. Cumulatively, the synergistic use of these antibody discovery methods yields 49 SARS-CoV-2 neutralizing antibodies against a variety of spike protein epitopes. Four antibodies and their binding mode were structurally characterized, revealing overlap with known antigenic supersites of the spike protein, and a unique quaternary binding mode at the receptor binding domain.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: Introduction

Antibodies serve indispensable roles in adaptive immunity by recognizing foreign molecular motifs. In this sense, antibodies are frontline soldiers against invading viruses, microbial pathogens, and their toxins[1]. However, antibodies also protect us against aberrant processes in our bodies by identifying abnormal molecular signatures of malignant growth and activating self-defense systems against cancerous cells. Antibodies are exclusively produced by B-cells and in theory, can be raised against any molecule, i.e., an antigen (antibody generator). Activated B-cells can be committed to immunological memory to be recalled later upon re-encountering their antigen. Their breadth of antigen specificity and capacity for high affinity binding makes antibodies attractive candidates in biotechnological applications and protein therapeutics[2]. Though the field of antibody discovery and engineering is advancing, isolating antibodies with specific binding characteristics remains a challenge in developing therapeutics. The work presented here addresses this bottleneck through engineering of a microbial antibody display strain and development of an abridged method for the identification of light chain partners for antigen-specific heavy chains.

1.1 Antibody Structure And Function

1.1.1 Immunoglobulin Fold

Antibodies are part of the immunoglobulin superfamily (IgSF), proteins possessing the signature immunoglobulin homology unit[3]. This evolutionarily ancient protein fold[4] is

often described as a flattened Greek key beta barrel because of the characteristic sandwich of two antiparallel beta stranded sheets at its core, stabilized by an intradomain disulfide bond (**Figure 1.1**)[5]. A full-length antibody consists of two heterodimers connected via disulfide bridges. One heterodimer in turn consists of a heavy and a light chain. Heavy and light chains are encoded separately and are linked to each other through a covalent disulfide bond. The heavy and light chains consist of consecutive globular Ig units similar to pearls on a string, each approximately 110 amino acids in length.

Antibodies have two types of immunoglobulin (Ig) units: a variable domain that mediates antigen binding and a constant domain that provides structural support and performs immune effector functions. The characteristic Y shape of antibodies couples their antigen recognition sites at the tips of the two Y arms to the immune effector domain that comprises the stalk. These domains are termed Fab (fragment antigen binding) and Fc (fragment crystallizable) referring to the arms and stalk of the Y, respectively (**see Figure 1.2**). Light chains only have one constant (C_L) and one variable (V_L) domain. Heavy chains also possess only one variable (V_H) domain but have three or four constant (C_{H1-4}) regions, depending on antibody class.

Light chains use either the Kappa (κ) or Lambda (λ) locus variable and constant genes, with roughly 60-65% of antibodies employing κ and 40-35% λ. Heavy and light chains covalently associate with each other through a disulfide bridge between the C_L domain and the first heavy chain constant region C_{H1} to form the arms of the Y shape[6]. Fabs consist of the entire light chain (one constant region C_L and the variable domain V_L) and two domains of the heavy chain: the C_{H1} and the heavy chain variable domain (V_H). The

C_L, C_{H1}, and V_H, V_L domains respectively share a binding interface with an extensive non-covalent interaction network[7].

The antigen-binding pocket is formed at the solvent-exposed tips of the V_H and V_L domains through peptide loops that connect individual beta strands of the barrel (**Figure 1.1**). The architecture of constant and variable regions differs in the number of beta strands comprising the two sheets. Constant domains total seven beta strands divided three to four between the bilayers, while variable domains have one additional shorter beta strand inserted on the edges of each sheet for a total of nine[8]. The loops connecting the beta strands are integral to how antibodies recognize and bind antigens. In variable domains, these loops are regions of hypervariable sequence space due to genetic diversity and rearrangement at the Ig locus. The loops comprise the complementarity-determining regions (CDRs), while the beta strands make up the framework regions (FR) and have considerably less sequence variation[9].

CDR variability determines the antigen specificity of a given antibody. Both heavy and light variable domains have three CDRs and four FRs each. Together, the heavy and light chain CDRs (CDRH, CDRL) create the antigen recognition site or paratope. Because six peptide loops contribute to the paratope, it forms complex three-dimensional landscapes that result in tailored binding crevices and/or protrusions to accommodate a wide range of shapes, sizes, charges, and hydrophobicity of potential antigens. CDRH3 often has the most extensive interactions with the antigen and is the major determinant of antigen specificity[10].

1.1.2 Processes Underlying Antibody Diversity

Antibody diversity is achieved through the segmented nature of the immunoglobulin heavy and light chain genomic loci, which requires recombination and joining of the appropriate antibody gene fragments needed to create a productive exon. Both heavy and light chain loci contain multiple variable genes arising from gene duplication events. Only one variable heavy and light gene can be expressed at a time[11]. The λ and κ immunoglobulin loci are split into variable, joining, and constant gene segments. The recombinases RAG1 and RAG2 recognize specific recombination signal sequences preceding variable and joining gene segments and excise the interspaced genomic DNA to enable joining of the coding segments and produce a full-length antibody construct[12]. This is an imprecise process that results in nucleotide deletions or additions in the V and J joint thereby increasing diversity[13]. The heavy chain locus is arranged similarly but has additional diversity elements that insert between the variable and joining segment. Diversity elements cause the increased diversity of the CDRH3[14].

This process is called V(D)J recombination and occurs during B-cell development in the bone marrow (**Figure 1.3**)[15]. The heavy chain locus of pre-B-cells is the first to rearrange and, only upon successful heavy chain creation, is rearrangement at the κ locus initiated. The λ locus only undergoes rearrangement, if productive κ light chain rearrangement fails. Pre-B-cells that cannot produce a rearranged heavy and light chain undergo apoptosis and never leave the bone marrow[16]. The result is a calculated possible diversity of 10^{11} distinct antibody sequences through recombination alone[9]. After B-cell maturation, naïve (i.e., non-antigen experienced) B-cells (nBCs) enter systemic

circulation while displaying their unique antibody on the cell surface as a membrane-bound receptor, the B-cell receptor (BCR). nBC activation initiates upon encountering an antigen matching their BCR paratope. nBCs can become either of two types of antibody-secreting cells: plasmablasts (PB) or plasma cells (PC)[17].

PBs arise before PCs in primary foci located near the medullary cords in lymph nodes 3-4 days after activation. To differentiate into PCs, activated B-cells (aBC) enter lymph node follicles and form germinal centers (GC) in which they undergo extensive rounds of somatic hypermutation (SHM) and class switch recombination (CSR)[18]. GC B-cells acquire an antibody secretory phenotype only after BCR affinity maturation is concluded. SHM is restricted to variable domains and mostly affects CDRs[19,20]. Activation-induced deaminase triggers nucleotide substitutions in the coding region through deamination of deoxycytidine, creating uracil[21]. In the process of repairing the U:G lesion, incompletely understood mechanisms that Franklin and Steele argue may involve RNA-directed DNA repair, lead to mutations of A:T base pairs[20,22]. This complex process further expands antibody diversity and demonstrates how antibodies can achieve exquisite specificity and superb affinity towards their antigens. Antibody applications as affinity reagents and, more importantly, as therapeutics are practically limitless because of their breadth of antigen recognition and distinguished binding ability.

1.1.3 Fc Effector Function

The Fc region was named for the ease with which it crystallized after an antibody was digested with the protease papain[23]. Functionally, the Fc domain mediates interactions

with the immune system, flagging the bound antigen to be targeted by the other arms of the immune system, such as complement activation or natural killer cell and macrophage activity through antibody-dependent cellular cytotoxicity (ADCC)[10]. Unbound, circulating antibodies do not trigger complement activation, as C1q requires binding of at least two Fcs simultaneously. Fc crosslinking occurs when antibodies are near each other, such as during the opsonization of antigens[24]. Depending on their use of one of five different heavy chains (α, δ, ε, γ, or μ), antibodies belong to one of five classes: IgA, IgD, IgE, IgG, or IgM. The subclasses differ in tissue distribution and the immune effector function of the Fc.

IgMs can be membrane bound or secreted, and as such are the class initially expressed by developing B-Cells as a cell-surface receptor (BCR). As the earliest antibodies, they usually possess inferior affinity; however, they can multimerize into pentamers and rarely hexamers through intermolecular disulfide bond formation between their Fcs. The crosslinking effectively increases their avidity and the pentameric structure potently activates complement[10]. IgD is co-expressed with IgM on the surface of B-cells and also exists in secreted form, although it constitutes less than 0.5% of circulating antibodies and has poor serum stability. Though its function is incompletely understood, there is evidence that IgD's significantly longer and more flexible hinge region connecting the Fab and Fc causes it to be exclusively responsive to polyvalent antigen interaction when in its membrane-bound form[25,26].

After activation, antibodies can undergo CSR in GCs, and switch which heavy chain they express in an antigen-dependent manner. IgE plays a prominent role in anti-parasitic

immunity and stimulates the activity of mast cells and basophils, evoking common symptoms of allergies through the hypersensitivity response[27]. IgGs are the most abundant immunoglobulin type in circulation, accounting for 75% of serum antibodies and are the most used therapeutic monoclonal antibody format[28]. IgGs can be further subdivided into G_1, G_2, G_3, and G_4, ordered by their prevalence in serum. Their effector functions differ subtly based on differences in their heavy chains. All four efficiently neutralize antigens and can mediate ADCC by binding one of the three Fcγ receptors (FcγR I, II, III); however, only three (G_1, G_2, G_3) can bind C1q to activate complement. Interestingly, exposure to allergens[29] and some parasitic infections[30] elicit an IgG_4 response that shows functional similarity to IgE. IgG_2 and, to a lesser extent, IgG_1 are elicited in response to polysaccharide antigens, while IgG_3, IgG_4, and IgG_1 preferentially arise against proteins[31].

IgA can exist in monomeric or dimeric form. Similar to IgG, IgA has three subclasses (IgA1, IgA2, and IgA3) that largely overlap in sequence. The monomeric form of IgA makes up 15% of circulating serum antibodies[10]. As a dimer, IgA is secreted at mucosal surfaces and is the main antibody class providing barrier protection. The J-chain, a 15 kDa protein, can form disulfide bonds between different Fcs of either IgM or IgA multimers, is required by IgA to dimerize, and is essential for IgA and IgM secretion[10].

1.2 Therapeutic Antibodies

For the average consumer, it is not immediately obvious how ubiquitous monoclonal antibody therapeutics are, especially in oncologic disease and autoimmune conditions.

Antibody drugs follow complicated naming conventions that result in tongue twister names such as Trastuzumab, Pembrolizumab, Ustekinumab and Adalimumab that sound foreign even to some patients currently receiving these life-saving interventions. However, their much more pronounceable brand names Herceptin, Keytruda, Stelara, and Humira are recognizable to us from prominent TV ads, their names quickly becoming as familiar as Tylenol or Advil. But rather than treat headaches and moderate pain, these biologics are used to treat breast cancer, melanoma, psoriasis, and rheumatoid arthritis (RA)[32]. In 2020, the pharmaceutical company Johnson and Johnson reported over $14 billion sales in their consumer health category, which includes Tylenol, Zyrtec, and Listerine. However, their most profitable sector was pharmaceutical, with over $45 billion sales, a growth of 8.4% that they contribute to their mAb therapeutics Stelara, Drazalex, and Erleada among other drugs[33]. Another statistic that illustrates the dominance of antibody therapeutics in the pharmaceutical drug market is Humira, which was developed by Abbvie and has been the world's best-selling drug since 2012, surpassing $20 billion in sales in 2020[34].

The first therapeutic antibody was approved by the US Food and Drug Administration (FDA) in 1986 and targeted the T-cell protein CD3. Muromonab was administered to kidney transplant patients to suppress xenograft rejection[32]. It took ten years to bring an antibody therapeutic to market after Kohler and Milstein successfully immortalized B-cells and created a hybridoma cell line capable of producing monoclonal antibodies (mAb) of known specificity[35]. Although many thought Paul Ehrlich's 'magic bullets'[36] were finally within their grasps – so strong was the conviction in their therapeutic success

– use of muromonab was limited as it was isolated from mice and often elicited an immune response in patients due to species incompatibility. It took over ten years for the first 'blockbuster' drug to enter the market. Still prescribed and widely used today, rituximab was approved in 1997 by the FDA[37] and targets the cell-surface protein CD20, which is overexpressed in B-cell malignancies[37]. Clinical trials demonstrated rituximab's effectiveness against non-Hodgkin lymphoma[38] and indolent lymphoma[39].

In the time between muromonab and rituximab development, Morrison et al. replaced mouse Fc and CH1 and CL domains with a human Fc region to create a chimeric antibody[40]. This technology was used for rituximab to prevent the immune reaction observed with muromonab. Rituximab grossed $5 million in its first year[41]. Since then, its sales have inflated to a high point of $8.58 billion in 2016[42] before rituximab patents expired. Subsequently, rituximab's annual revenue decreased to $4.52 billion in 2020 because of biosimilars entering the market at lower costs[43]. However, even for chimeric antibodies, the human anti-mouse antibody (HAMA) immune response to murine sequences in the variable domain can occur. Winter et al. addressed this issue by humanizing the variable region through grafting of murine CDRs onto human framework regions[44]. In 2020, 69.4% of approved mAbs of non-human origin had been humanized and 25% were chimeras[32,45].

Adalimumab (Humira) was the first fully human mAb to be approved by the FDA in 2002 and was discovered through phage display rather than hybridoma screening[46]. This technological advance emancipated antibody discovery campaigns from the limitation of mining mice repertoires, and human mAbs now account for the largest segment of

approved therapeutics[32,45]. Continued innovation over the last twenty years has expanded the range of antibody treatment modalities and introduced novel therapeutics, such as drug-conjugated mAbs and bispecific antibodies.

For over thirty years, the number of therapeutic antibodies has steadily increased and according to The Antibody Society, 135 have been approved either in the US or EU[45] as of February 17, 2022. By far the largest indication group is in oncology, with 45% of antibodies targeting cancer-related antigens, followed by immune disease targets at 27% market share. In 2020 the global antibody market reached $157.33 billion, and it is projected to crack the $500 billion mark by 2030[47,48].

1.3 Antibody Discovery Techniques

Antibodies were first discovered in 1890 by Behring and Kitasato as the substance in the serum of diphtheria and tetanus-exposed animals that could neutralize the toxins[1]. Their contemporary Paul Ehrlich realized what enormous potential antibodies held as a highly targeted therapeutic that would spare surrounding healthy tissue early in his research[36]. However, isolating and producing a specific antibody of interest from the polyclonal repertoire of billions of B-cells circulating in the human body[49] was a technological challenge that needed to be overcome first.

1.3.1 B-Cell Hybridomas

Kohler and Milstein's 1975 landmark paper on the creation of B-cell hybridomas strongly impacted antibody therapeutics development. Kohler and Milstein exploited the fact that myelomas and cancerous plasma cells secrete clonal antibodies of singular specificity and

possess a proliferative phenotype, in contrast to short-lived B-cells. They fused splenocytes from mice immunized against sheep red blood cells with a murine myeloma and selected for successfully hybridized cells. These hybridomas maintained the proliferative phenotype of the myeloma and obtained antigen specificity against sheep red blood cells, essentially creating an immortalized antibody-secreting cell of known specificity[35,50].

While revolutionary, the hybridoma technique was not without–drawbacks. In the beginning, its utility was limited to the discovery of mouse antibodies because no suitable myeloma cell lines from other animals were available and inter-species fusion of myeloma and B-cells (hetero-hybridoma) led to chromosomal instability[51]. Within time, myeloma cell lines for common species, such as rats, chickens, rabbits and humans were developed[51]. However, a human hybridoma that fulfills the promise of simple human therapeutic antibody discovery has remained elusive for multiple reasons. For one, the fusion efficiency of human B-cells is extremely poor at a rate of 1:100,000 cells fusing successfully[52]. Additionally, procuring a sufficient amount of antigen-specific B-cells is challenging as they are observed in peripheral blood infrequently[53], and ethical concerns preclude extraction of antigen-experienced B-cell enriched tissues, such as the spleen or lymph nodes. Similarly, immunization of humans with antigens to elicit antibodies of interest is limited.

Hybridoma technology relies on the naturally produced antibody repertoire of the immunized animal which comes with advantages and drawbacks. On the one hand, natural antibody heavy and light chain pairs are maintained during fusion, as well as the

mAb's specific Fc subclass, which is driven by characteristics of the antigen during CSR[54]. On the other hand, ~~it~~ the technology is not suitable for poorly immunogenic antigens that do not elicit a strong immune response, which includes some prevailing pathogens[55]. Additionally, immune tolerance prohibits raising antibodies against self-antigens, as self-reactive B-cells are terminated in the bone marrow before entering circulation. While this protective mechanism prevents autoimmunity, many therapeutically relevant mAbs in oncology and immune disorders target endogenous self-antigens, such as PD1 (programmed death receptor 1). Cancer cells express PD1 on the surface to suppress T-cell activity directed against them[56]. Anti-PD1 antibodies like nivolumab[57] bind PD1 and prevent it from engaging T-cells, enabling them to recognize and destroy the cancerous cell. Finally, antibody discovery using hybridoma generation is a time and resource-extensive enterprise. From animal immunization and hybridoma generation to screening of successfully fused cells for candidates, the typical timeframe is weeks to months and demands the maintenance of large cell banks and care of laboratory animals.

1.3.2 Surface Display Technologies

The crux of antibody discovery is the maintenance of genotype to phenotype linkage (i.e., antibody sequence to specific antigen binding). This is a bottleneck especially for functional repertoire analysis and high-throughput efforts, which aim to interrogate upwards of millions of antibodies in an efficient and cost-effective manner. To overcome this obstacle, antibodies can be displayed on the surfaces of phage, yeast, or mammalian

cells to select specific antigen binders. An advantage of this technique is that all members of an antibody library mined from immune repertoires can be interrogated for antigen-specific binding at the same time.

Greg P. Smith pioneered phage display technology in 1985 when he published the successful expression of peptides on the surface of bacteriophage M13's coat proteins[58]. Following his discovery, three independent groups demonstrated antibody fragment display by phages. They used a fragment consisting of covalently linked variable regions of an antibody called single-chain variable fragment (scFv)[59–61]. The reduced format enabled more efficient surface display as it lessened the protein folding burden of the non-endogenous antibody structure for bacteria.

Phage display produces antibodies from recombinant libraries which are then enriched for binders of a specific antigen through several rounds of in vitro biopanning. Original phage display library sizes can be up to 10^{11}, allowing for much higher throughput than hybridoma screening. The source of the libraries can be B-cell repertoires of immunized animals or humans, similar to hybridomas or can be antigen-agnostic by employing naïve repertoires or entirely synthetic libraries[32,62,63]. Using naïve or synthetic libraries overcomes the limitations of naturally occurring immunity in response to immunization, and self-reactive clones are not pruned from the repertoire of synthetic libraries. Adalimumab (Humira), the world's best-selling drug nine years running with sales of $20.1 billion in 2020[34], was isolated from a human phage display library mined for binders of tumor necrosis factor-alpha (TNF-α), an endogenous human protein[64]. TNF-α is a proinflammatory cytokine that can exist in membrane-bound form[65] and is a

therapeutic target in many auto-inflammatory diseases, such as rheumatoid arthritis, psoriasis, and Crohn's disease[66–68]. Adalimumab can neutralize TNF-α in both its soluble and membrane-bound form[64].

Immune libraries can be obtained from the reverse transcription of peripheral blood mononuclear cells (PBMCs) mRNA, where the immunization state of the donor determines whether the library is antigen polarized. If naïve libraries are desired, the RT can be performed with IgM-specific primers to avoid amplification of antigen-experienced IgG or other class-switched subclasses. Both scFv and Fab antibody libraries are commercially available and have demonstrated that their repertoires are sufficiently diverse to yield high-affinity binders to a wide range of targets[69,70].

To date, nine antibodies from phage display discovery campaigns have received FDA approval[45]. Since the conception of phage display, other display modalities have been developed, including ribosomal display, yeast display, and mammalian display[71–73]. These all have application-specific advantages and drawbacks for antibody discovery and high-throughput repertoire interrogation, which I will be review in-depth in Chapter 2.

1.3.3 Transgenic Animals

At the time of their development in the second half of the 1990s, transgenic mice filled the crucial gap in hybridoma technology between the need for human-derived antibodies and the lack of a suitable human myeloma cell line for B-cell fusion[74]. The two transgenic mice models HuMabMouse[75] and XenoMouse[76] were engineered to exclusively express human antibody sequences. Human IgH, IgL, and IgK sequences

were introduced into mouse embryos on yeast artificial chromosomes and the resulting mice were crossbred with immunoglobulin-deficient strains[77]. These strains developed B-cells normally and upon immunization, mice would produce antigen-specific, fully human antibodies[75,78–80]. In 2006, the first XenoMouse-derived therapeutic antibody was approved for the treatment of colorectal cancer[81]. Panitumumab binds epidermal growth factor receptors and was the first mAb to target solid tumors[82]. Another prominent therapeutic, Stelara or ustekinumab, was discovered through the HuMabMouse and targets the p40 subunit of the cytokines interleukin 12 and 23. The immunosuppressive result has demonstrated effectiveness in the treatment of psoriasis, ulcerative colitis, and Crohn's disease[83–85].

One drawback of the XenoMouse and HuMabMouse is the limited ability of the human constant regions of the immune effector domains to be recognized and interact with the mouse immune system to undergo SHM and CSR. To overcome this, second-generation transgenic mice and rats that maintain their endogenous constant heavy chains while possessing all human variable immunoglobulin genes were created[86,87]. Recently, these second-generation mice strains demonstrated their powerful antibody discovery potential when the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) swept across the globe[88]. Baum et al. expeditiously launched an antibody discovery campaign using Regeneron's VelocImmune transgenic mouse model and identified virus-neutralizing candidate mAbs[89,90]. REGN10987 and REGN10933 were selected for advancement to clinical testing[91] as a cocktail and received FDA emergency use authorization on the 21st November of 2021[92].

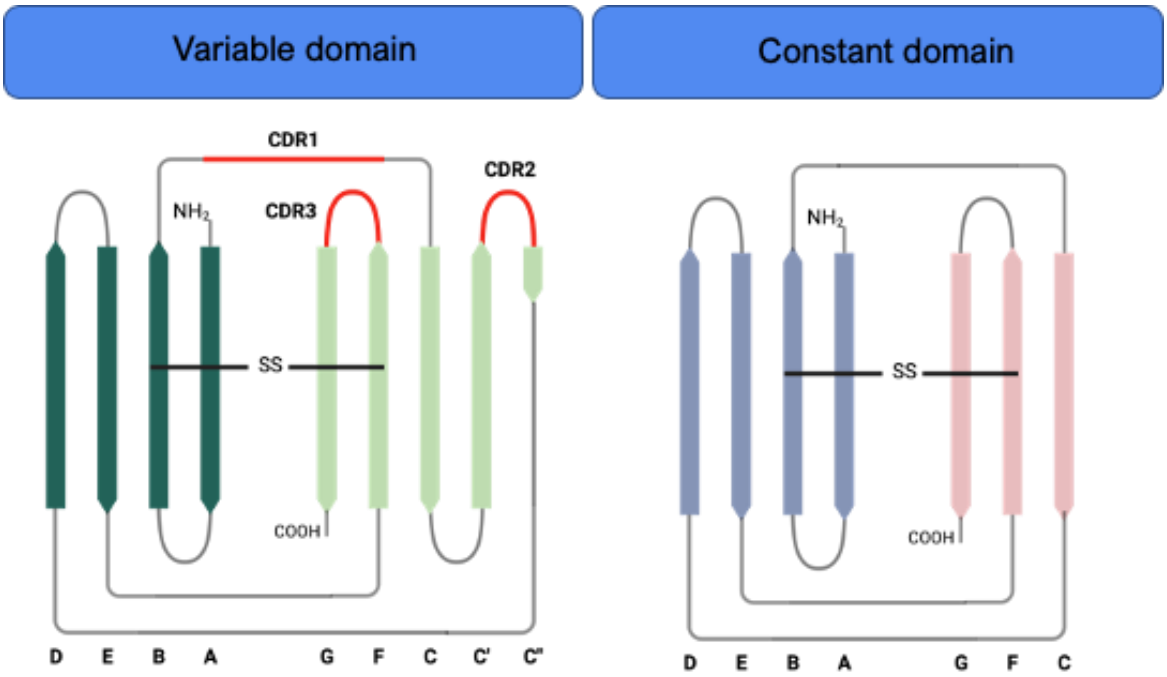

Figure 1.1. The Greek key barrel fold of immunoglobulins. The immunoglobulin fold consists of a conserved motif of two beta-strand sheets forming a flattened barrel structure around an inter-strand disulfide bond. The variable domain has two additional beta-strands compared to the constant domain that are of functional significance. The extra loop connecting C' and C'' forms the CDR2 and is part of the paratope. This figure was produced using BioRender and is printed with permission. Created with BioRender.com

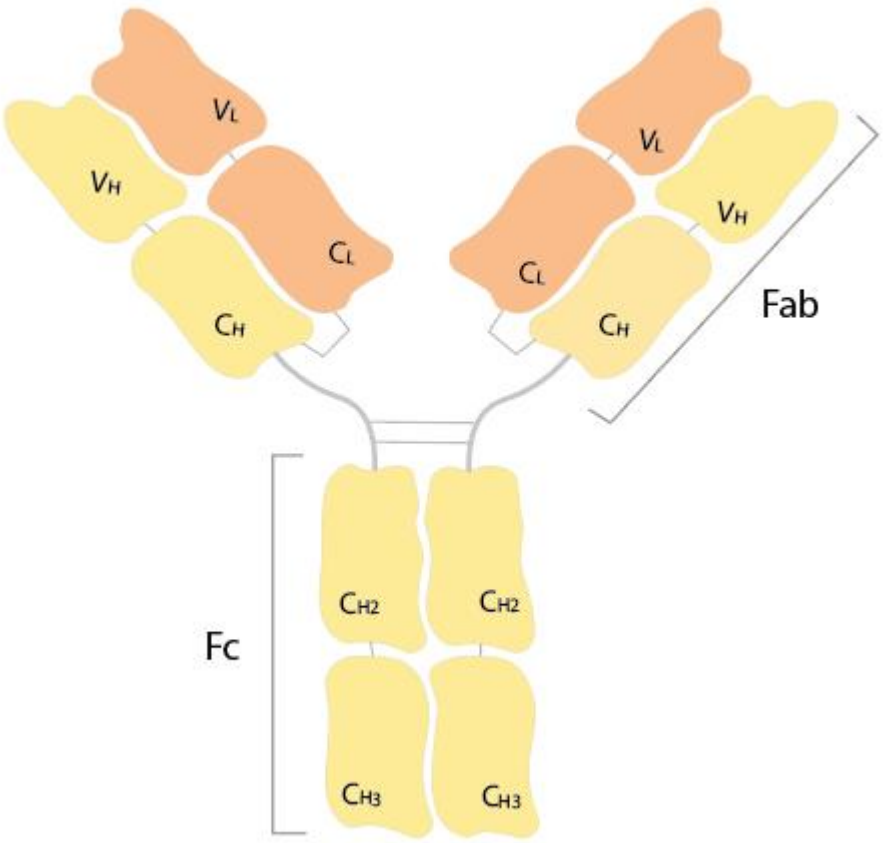

Figure 1.2. Structure of immunoglobulins. The full-length antibody is Y-shaped, the stalk region (Fc for fragment crystallizable) consists entirely of the heavy chain and only varies in constant domain composition between immunoglobulin subtypes. The 'branches', antigen-binding fragments (Fabs), contain the antigen-specific variable domains at their solvent-exposed tips. Fabs consist of the entire light chain (one constant region C_L and the variable domain V_L) and the two domains of the heavy chain it shares binding interfaces with heavy chain constant domain 1 (C_{H1}) and the heavy chain variable domain (V_H). C_{H1} and C_L are the domains connected via a disulfide bridge[6].

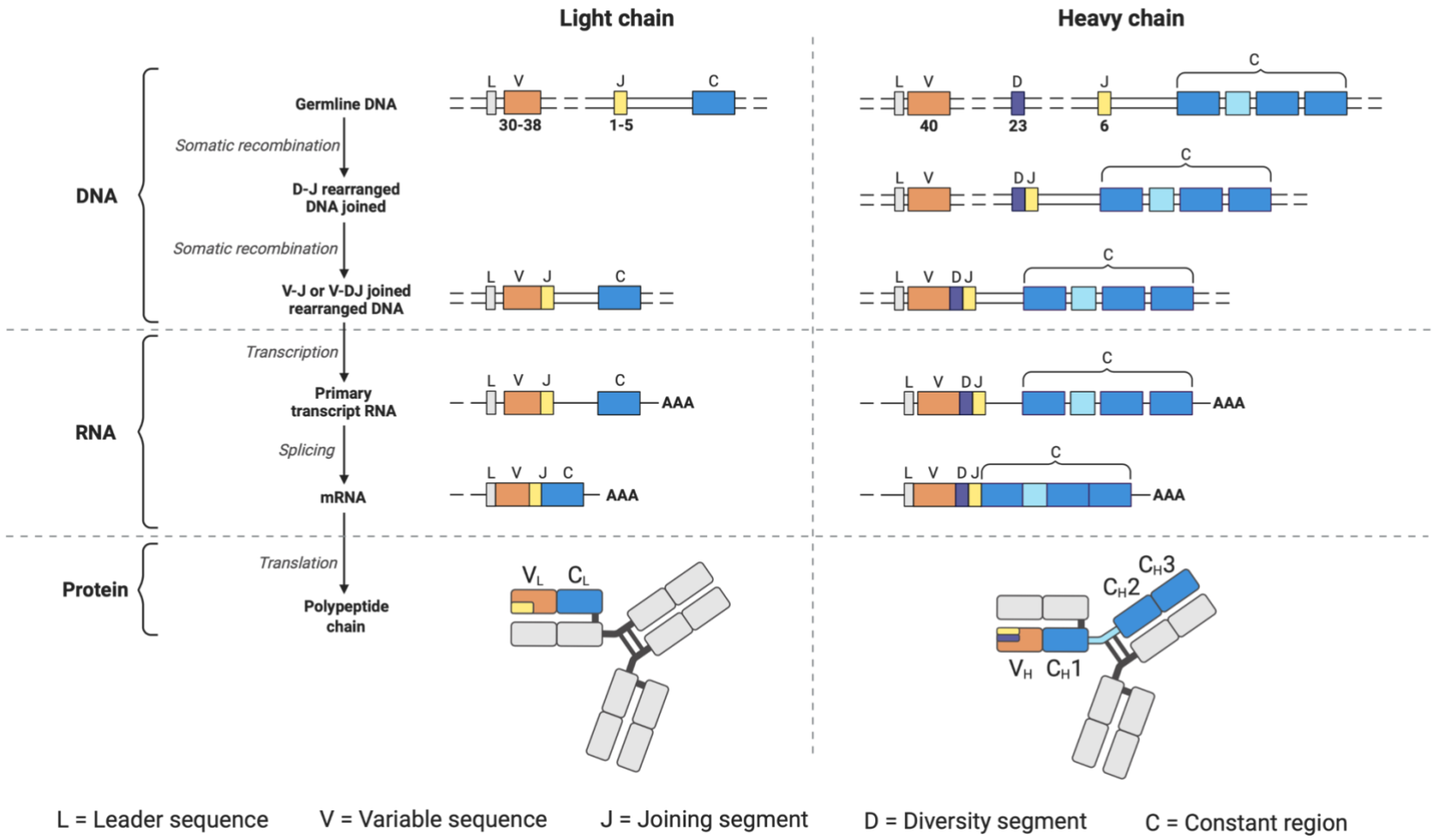

Figure 1.3. V(D)J recombination is the process underlying antibody diversity. V(D)J recombination occurs in both the heavy and light chain loci. Joining of the different gene segments that comprise the full-length variable region follows a sequential process. In heavy chains, the D and J segments are rearranged first, followed by combination with a V-gene before the constant domain is added. At the light chain locus the diversity element is absence and the first rearrangement step is the combination of a variable and joining gene segment. This figure was produced using BioRender and is printed with permission. Created with BioRender.com

CHAPTER 2: Engineering of *S. cerevisiae* for Fab display

2.1 INTRODUCTION

2.1.1 Antibody display technologies

This chapter describes the development of a high-throughput antibody discovery platform as part of a rapid response initiative for the defense against emerging microbial threats. Traditional methods, such as hybridoma screening and single B-cell sorting, are relatively low throughput and time intensive[51]. Antibody display methods (most notably phage display) can mine larger libraries than single-cell-based approaches and have a proven track record of therapeutic antibody discovery (e.g., Humira)[70,93]. However, every display technology comes with drawbacks and advantages.

The main restriction for display technologies is the choice of antibody display format. Antibody folding is a multi-step process requiring specific chaperones and the specialized redox environment of the endoplasmic reticulum (ER) for correct formation of disulfide bonds[94]. Intra-disulfide bonds stabilize each Ig-domain's beta barrel core, and inter-disulfide bonds provide the Fab heavy and light chains with their only covalent linker[10]. Bacteria that carry the protein folding burden in phage display lack sophisticated eukaryotic protein quality control mechanisms and have cytosolic environments that impede efficient disulfide bond generation because they lack organelles, including an ER.

While yeast maintains a protein folding environment more closely related to mammalian cells[95], antibodies, non-native proteins containing a structural fold found only in higher animals[4], burden its translational system. To mitigate this folding stress Fab

fragments, the smallest stable and native antibody format, are often altered into single chain variable fragments (scFvs). ScFvs are comprised of heavy and light variable domains held together by an artificial linker sequence (**Figure 2.1**). Wittrup and Boder pioneered yeast display using scFvs[72]. Since then, yeast display has been used for antibody affinity maturation, characterization and isolation of binders from immune and synthetic libraries[96–98]. However, traditional yeast display platforms are restricted to non-native antibody constructs for efficient display, which diminishes their convenience, versatility, and efficiency.

ScFvs often differ substantially in native binding efficiencies and biophysical properties when reformatted into Fab fragments or full-length antibodies[99–101]. When comparing the neutralizing ability of four scFvs against the Cn2 component in scorpion venom, Hernandez et al. found that three became neutralizing when expressed as Fab fragments in contrast to the scFv. Additionally, the Fab fragments retained binding ability at higher concentrations of guanidinium chloride compared to their scFv counterparts, suggesting increased stability[100]. Thus, the use of scFvs distorts analyses of immune repertoires and complicates downstream antibody developability.

Current yeast display formats have limited ability to display Fab fragments and must contend with high numbers of non-expressing cells when Fabs are used, leading to an underrepresentation of repertoire diversity. However, even under these inadequate assay conditions, Fabs enable more accurate characterization of binding efficiency

compared to scFvs and other artificially linked constructs[101], in addition to being a more relevant format for therapeutic antibody design[102].

Though mammalian cells are better suited to produce and display antibodies, mammalian surface display requires time-consuming and laborious transfection protocols and is limited to library sizes of ~10^6. Library transfection is often only transiently achieved because of the low efficiency of stable integration and associated complications, leading to loss of library members over time[103,104]. These factors make mammalian display of antibodies unsuitable for high-throughput, routine screening of repertoires and iterative interrogation of heterogeneous antigen-binding profiles.

Considering the limitations of each technique, we selected yeast as the host organism of a rapid antibody display platform. This platform needed to leverage high-throughput repertoire sequencing and efficient genotype-to-phenotype linkage for the facile screening, interrogation, and characterization of antibody repertoires en masse. We required a more sophisticated and robust yeast display strain capable of reliable, reproducible expression of correctly folded Fabs for accelerated antibody discovery and comprehensive repertoire interrogation. To this end, we hypothesized that a humanized yeast strain closely mimicking the native folding environment of antibodies in antibody-secreting plasma cells could improve antibody folding through interactions with cognate chaperone proteins.

A 1998 study used protein disulfide isomerase (PDI) and binding immunoglobulin protein (BiP) chaperone co-expression to increase scFv secretory capacity in *S. cerevisiae*

up to 8-fold[105]. In the context of our current understanding of antibody folding, the observed increase is likely attributable to a global relief of folding stress through the upregulation of chaperones, rather than the specific contribution of chaperones to scFv folding. Since these first experiments over 20 years ago, antibody folding has been extensively studied and is now understood in much greater detail[7,94]. Cumulatively, this work demonstrates that antibody folding is intricately linked to specific chaperone interactions at distinct times during the hierarchical folding cascade (**Figure 2.2**). Because C_{H1} subunit folding (not present in scFvs) functions as a crucial quality control mechanism in antibody heterodimer assembly[106], we focused its productive folding at the center of our design efforts.

2.1.2 Introduction: Antibody Folding

When B-cells differentiate into antibody-secreting PCs, a transcriptional network is activated to ensure simultaneous expression of chaperones necessary for antibody folding. Figure 2 outlines the folding pathway of Fabs and their interacting chaperones. The main obstacle in protein folding is the intrinsically unfolded nature of the constant domain 1 (C_{H1}) of the heavy chain[106]. BiP is required to bind the C_{H1} subunit in the absence of its native binding partner, the constant region of the light chain (C_L)[107]. While bound to BiP, C_{H1} is oxidized and forms the intrachain disulfide bond characteristic of the immunoglobulin fold. PDIs are instrumental in catalyzing these reactions and directly interact with antibodies in crosslinking studies[108] and act cooperatively with BiP in the oxidation of the intradomain disulfide bonds[109].

Further folding requires binding of C_{H1} to the fully folded C_L domain of the light chain. Association of the C_{H1} folding intermediate with C_L is fast, however, its subsequent collapse into its native fold and interchain disulfide formation between C_L and C_{H1} critically depend on a prior trans-to-cis bond conversion that changes the orientation of a conserved proline residue[106]. This rate-limiting step is catalyzed by peptidyl-prolyl isomerases (PPIs)[110]. Notably, chaperones do not act in isolation but as part of a finely tuned assembly line to create a correctly folded antibody. Additionally, cofactors are instrumental in maintaining chaperone function. Specific chaperones selected based on the outlined rationale are summarized in **Table 2.1**.

To compensate for the increased translational burden in human PCs, the unfolded protein response (UPR) regulators IRE1α and XBP1 trigger expansion of the ER[111]. For our purpose of producing high levels of correctly folded, heterologous protein in *S. cerevisiae*, activation of the UPR can be a double-edged sword. On the one hand, alternative splicing of the *S. cerevisiae* transcription factor HAC1p by ER-stress sensing protein Ire1p triggers increased transcription of chaperones under control of the UPR element[112,113]. This alleviates the burden of unfolded proteins in the ER – a useful mechanism to cope with the additional translational demands imposed by heterologous protein expression. However, on the other hand, its sustained activation leads to cell cycle arrest and ultimately apoptosis[114].

An *S. cerevisiae* knockout strain lacking the negative lipid synthesis regulator OPI1 has been shown to cause ER expansion and increase yield of heterologous IgG up

to four-fold[115,116]. Based on this evidence, the careful calibration of chaperone transcription by HAC1p and ER expansion through deletion of OPI1 may improve folding of Fabs in *S. cerevisiae*.

2.2 RESULTS

2.2.1 Genomically integrated Fabs display homogenously

To enable facile and rapid cloning of antibody libraries, we designed our antibody display vectors to be compatible with the modular synthetic biology workflow outlined in the Yeast Tool Kit (YTK)[117]. Heavy and light chain V-genes are amplified from cDNA, and AaRI restriction sites are added for Golden Gate[118]-based assembly into a custom dropout (DO) vector (**Figure 2.3 A**). The DO vector encodes red fluorescent protein (RFP) and green fluorescent protein (GFP) as placeholders for the variable heavy and variable light chain, respectively. Upon AaRi digestion, RFP and GFP genes are excised, and the variable inserts ligated into the vector. Screening for the absence of red or green signals enables simple color-based selection of successful transformants.

To preserve the native antibody binding interface as much as possible, we chose structurally unaltered Fabs as our display format. Therefore, the variable inserts form a scarless transcript with either the CH1 or CK/CL domain after ligation into the DO vectors. Transcription is driven from a bidirectional, gal-inducible promoter. The bidirectional promoter allows expression of the Fab's heavy and light chain from one construct, and its galactose-inducible nature enables temporal control of Fab expression. To ensure homogeneous expression, the antibody display vectors integrate into the yeast

genome through homologous recombination. Genomic integration also avoids plasmid loss during cell division over time and uptake of multiple plasmids simultaneously. We selected the *LEU2* locus for integration of the antibody display cassette, which also includes AGA2 to anchor the Fabs to the yeast cell wall (**Figure 2.3 B**)[72].

The light chain contains an N-terminal FLAG sequence that is used to monitor Fab expression through fluorescent labeling with fluorophore-conjugated affinity reagents. The membrane anchor AGA2 is translated in a continuous transcript with the heavy chain. Therefore, binding and expression signals will not co-localize if the CH1 and CL domains fail to form interchain disulfide bonds. When integrated, 99.14% of the test expressor P4b3 cell population expressed Fab above the background level of autofluorescence, displaying a homogenous shift in the population with very few non-expressing cells (**Figure 2.3 C**).

2.2.2 PDI and BiP, but not BiP alone, improve Fab expression

Fabs are a preferable display format for antibody discovery because Fabs maintain the native VH:VL configuration and are smaller than full-length IgGs, which reduces the protein folding burden for the display organism. To improve Fab folding and expression in the yeast display strain EBY100, we engineered yeast ER to optimize heterologous protein expression. We identified the chaperones in **Table 2.1** as functionally significant in the antibody folding pathway. We constructed proof-of-concept chaperone strains and tested the expression of three antibodies: Ma6, Ma580, and Ma663. The three antibodies intrinsically differ in expression levels in the order of Ma6 >

Ma663 > Ma580 (**Figure 2.4 A**). This makes the three mAbs suitable tools for evaluating changes in expression levels over a varied range of intrinsic expression ability.

Initial strains were constructed in wild type (WT) EBY100 strains containing either BiP, PDI, a non-coding spacer, or PDI and BiP (P&B) simultaneously. Both PDI and P&B improved Ma6 and Ma663 Fab expression over spacer levels (**Figure 2.4 B, C**). BiP alone, however, had no positive effect on expression and slightly reduced expression levels of Ma663 compared to spacer. The lack of effect was not unexpected, as BiP's known role as a holdase in the antibody folding process might cause Fabs to remain sequestered in the ER[94].

Ma6, the best expressor, showed no significant difference in expression in the PDI versus P&B strains; however, both strains improved expression of Ma6 levels compared to the spacer strain. The medium expressor Ma663 showed a discrete improvement in Fab expression in the double chaperone strain P&B versus PDI. No clear trend for expression improvement was detected for the weakest expressor, Ma580. Ma580 intrinsic expression levels are only weakly above background fluorescence and are not significantly influenced by the presence of chaperones (**Figure 2.4 D**).

Ma580 and Ma663 were isolated by Wardemann et al. from donors after natural infection with the malaria parasite plasmodium falciparum[119]. Ma6 was discovered from a clinical study participant receiving the R21 malaria vaccine (Ippolito, personal communication). R21 is based on the circumsporozoite coat protein (CSP) of the parasite's sporozoite life cycle stage, which is characterized by repeating motifs of the

amino acid (a.a.) sequence NANP[120]. Ma580 and Ma663 were discovered through single cell sorting of B-cells binding the CSP and subsequent crystallography studies mapped both mAbs' epitope to NANP repeats, albeit in different structural conformations.

However, when we tested binding of surface-displayed Ma580 and Ma663 Fabs to a biotinylated peptide containing six repeats of NANP ($NANP_6$), we detected no binding (**Figure 2.5 A**, **C**). We also tested Ma580 and Ma663 binding to a nineteen a.a peptide comprised of the sequence stretch connecting the CSP N-terminal domain to the central NANP repeat domain. This junctional peptide (JP) contains a singular NANP motif and did show binding to Ma663. Binding -concentration curves estimated a K_D of 6.5 μM for Ma663 : JP. Wardemann et al. report a K_D of 14.8 nM for Ma663 (IgG) : $NANP_5$[119]. The inability of surface-displayed Ma580 to bind neither $NANP_6$ nor JP is somewhat unsurprising considering the low levels of expression and fast k_{off} rate reported by Wardemann et al[119]. Surface-displayed Ma6 binds $NANP_6$ but not JP (**Figure 2.5 B, D**). Yeast display-based concentration curves estimated a K_D of 3.5 nM that aligns with EC50 values of 1.2 nM from ELISA testing reported for full-length Ma6 IgGs (Ippolito, personal communication).

2.2.3 BY4741 displays Fab fragments and reduces the induction time by one-day

Previous studies demonstrated that deletion of *OPI1* increased secretion of IgG in the W303 *S. cerevisiae* strain[121]. We decided to test whether deletion of *OPI1* in EBY100 would positively affect levels of Fab display. Initial data comparing expression of Ma580 in *OPI1* WT and Δ*opi1* were promising (**Figure 2.6 A**). However, while Δ*opi1* showed

normal growth, the cells acquired a flocculation phenotype that complicated the handling of its liquid culture and flow cytometry assays as cells were rapidly clumping together and settling in the tube (**Figure 2.6 B, C**).

In addition to the undesirable flocculation phenotype under Δ*opi1*, EBY100 revealed incompatibilities with the YTK design framework. Of foremost concern was the apparent endogenous expression of the endonuclease I-SceI. Duber et al demonstrated that prior introduction of the I-SceI recognition sequence in the location targeted for genomic integration increased integration efficiency when constructs were co-transformed with a cutter plasmid encoding I-SceI[117]. When we introduced I-SceI recognition sites at antibody integration loci in the EBY100 genome, transformed clones displayed unusual growth lag phases when inoculated into YPD (data not shown). As I-SceI is encoded by a mitochondrial intron in *S. cerevisiae*[122], endogenous expression of I-SceI in EBY100 and subsequent introduction of double stranded breaks at its recognition sequence site are the likely cause of the slowed growth during DNA repair. Unwanted I-SceI activity is absent in other *S. cerevisiae* strains, such as BY4741, due to genetic differences in the mitochondrial genome. Additionally, BY4741 carries a deletion for the flocculation protein 1 that may mitigate Δ*opi1* induced flocculation. For these reasons, we decided to create a BY4741-based display strain.

We based our design on the AGA1-AGA2 anchor system described by Boder et al[72], and integrated an extra copy of AGA1 under a galactose-inducible promoter at the URA locus of BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0). To mimic the parent

strain of EBY100, BJ5456[123], we deleted the proteases Prb1 and Pep4, as well as Bar1[124]. To test the function of this new BY4741-based display strain (BYDis), the antibody Fab P4c3 was integrated into the LEU2 locus of BYDis and EBY100 (WT). Expression and binding were compared for one-day and two-day induction periods at three temperatures: 30°C, 25°C, and 20°C.

Analysis revealed that a one-day induction period caused greater P4c3 Fab expression and binding in BYDis compared to a two-day induction period across all temperatures tested (**Figure 2.7 A-C**). The best expression and binding were observed for BYDis cultures induced and grown at 25°C for one day. Cultures induced for two days showed overall decreased expression and binding compared to one-day induction. At two days of induction, increased expression and binding correlated inversely with temperature: 20°C showed the greatest amount of expression and binding and 30°C the least (**Figure 2.7 D-E**).

When compared to EBY100, BYDis expression and binding were similar, though generally lower at the equivalent conditions tested, with the exception of a one-day induction at 20°C where BYDis performed marginally better than EBY100 (**Figure 2.7 D-F**). However, with one-day induction, BYDis had fewer non-expressing cells and a more homogeneously displaying population compared to EBY100. Up to 17% of EBY100 cells were non-expressing at 20°C. At the best display condition for BYDis (25°C for one day), an average of 1.6% of cells were non-expressing compared to 8.5% of EBY100 cells at the same condition (**FIGURE 2.8 A-C**). To confirm the display

behavior of BYDis, a second antibody Fab, P4b3, was integrated and tested through the same induction conditions. P4b3 reached higher levels of expression and binding compared to P4c3 but followed the same trends of expression and binding: one-day induction was preferable, and 25°C was the best expression temperature (**Figure 2.8 D, E**).

After establishing that BYDis satisfactorily displays antibody Fab fragments, we tested whether display could be improved by deletion of *OPI1* and still avoid a flocculation phenotype. Encouragingly, after CRISPR-Cas-mediated knockout, BYDis Δ*opi1* cultures did not flocculate or show otherwise abnormal growth behavior. We proceeded to integrate P4b3 and P4c3 into the BYDis Δ*opi1* strain and compared binding and expression to BYDis WT. BYDis Δ*opi1* displayed more P4b3 Fabs and improved binding. While the increase in Fab expression was not statistically significant, binding showed significant improvement: from 1459 RFU to 3197 RFU (**Figure 2.9 A, B**). P4c3 in BYDis Δ*opi1* did not show a clear improvement. P4c3 expression levels in Δ*opi1* were highly similar to WT and binding was slightly, but not significantly, decreased in Δ*opi1* compared to WT (**Figure 2.9 C, D**).

As integration of chaperones integral to antibody folding showed improvement in antibody expression levels in EBY100, we pursued this strategy to enhance BYDIs Δ*opi1* antibody display. We constructed two libraries of chaperones with three different promotor strengths in order to determine the best chaperone combination and expression strength for antibody display through flow cytometry-based cell sorting. One library

expressed human chaperones while the other used their yeast homologues instead. **Table 2.1** depicts the chaperones in each library. Additionally, a non-coding spacer was included in the libraries at each chaperone position as a negative control. The libraries were integrated into BYDis Δ*opi1* containing either Ma6 or Ma580. Selection for increased expression and binding did give rise to an AF-647+ population (**Figure 2.9 E**); however, AF-647 staining in the absence of antigen showed an undistinguishable increase in AF-647 signal. Therefore, the AF-647+ population cannot be attributed to improved binding (**Figure 2.9 F**). Single clones derived from AF-647+ cells showed discrete modes of fluorescence within the same population. This heterogeneous cell behavior did not seem linked to cell size or granularity, as highly AF-647 positive cells' scatter patterns were consistent with that of typical fluorescent cells of the population (**Figure 2.10**).

2.3 Materials and Methods

Strains and media. *E. coli* strain DH10B was used for all routine cloning of yeast and mammalian constructs. *S. cerevisiae* strain EBY100 (MATa AGA1::GAL1-AGA1::URA3 ura3-52 trp1 leu2-delta200 his3-delta200 pep4::HIS3 prbd1.6R can1 GAL) was acquired from ATCC (cat. no. MYA-4941) and initially used for antibody expression and selection. BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) was engineered to display antibodies. BY4741 was a kind gift from Dr. Elizabeth Gardner. Human chaperones were acquired as gBlocks from IDT. Yeast orthologue chaperones were amplified from yeast genome extractions of EBY100 with gene-specific primers.

The four EBY100 strains created were PDI, BiP, PDI and BiP, and a non-coding spacer sequence (YTK part 048)[117]. All chaperones were genomically integrated as an expression cassette in the HO locus under constitutive promoters. Routine culture was performed using YPD (Takara, cat. no. 630409). Antibody transformants were selected based on leucine prototrophy after genome integration (Takara cat. no. 630310). HO locus integrations were selected based on histidine prototrophy. Antibody expression was induced in YEP-galactose (1% yeast extract, 1% bacto-peptone, 0.5% NaCl, 2% galactose, 0.2% glucose).

Antigens and antibodies. Biotinylated JP (biotin–KQPADGNPDPNANPNVDPN) and $NANP_6$ (biotin-NANPNANPNANPNANPNANPNANP) were obtained from ABclonal and labeled with streptavidin-AF647 (Invitrogen, cat. no. S32357). Recombinantly expressed D614G contained a StrepTag II that was used for labeling by a rabbit anti-StrepTag II antibody (Biovision Cat. no. 3988-100) and a secondary label of goat anti-rabbit AF-647 (Fisher Scientific Cat. A27040). Fabs were labeled with anti-FLAG M2-FITC (Sigma, cat. no. F4049) or mouse anti-FLAG (Sigma Aldrich cat. no. F3165-1MG) with goat anti-mouse AF-488 (Fisher Scientific cat. no. A28175).

Antibody assembly and bacterial transformation. Antibody VH and VL sequences were amplified from Twist vectors with AaRI overhangs for assembly into custom antibody display vectors. Golden Gate assemblies were prepared in 20 μL reactions as follows: 2 μL 10X AARI buffer (Thermo Fisher Scientific, cat. no. B27), 0.4

μL 50X oligo buffer (ThermoFisher Scientific, cat. no. ER1582), 0.2 μL 100 mM ATP (ThermoFisher Scientific, cat. no. R0441), 20 fmol backbone DNA, 40 fmol VH and VL amplicons, 0.5 μL (2 U/μL) AaRI endonuclease (ThermoFisher Scientific, cat. no. ER1582), and 0.5 μL T7 ligase (3000 U/μL) (NEB, cat. no. M0318). Thermocycling protocol was: 37°C, 2 min, 16°C, 1 min; go to step 1, x29; 37°C, 30 min; 80°C, 15 min; hold at 4°C. Routine cloning was performed using Mix and Go (Zymo Research T3001) competent DH10Bs.

Transformation into yeast and protein expression. Purified plasmids were linearized for integration into the yeast genome via homologous recombination at the LEU2 or HO locus. For each 1 μg of plasmid, 0.5 μL NotI (10 units/μL) (NEB, cat. no. R0189) was used with the supplied buffer 3.1 in 10 μL. Reactions were incubated at 37°C for a minimum of 5 hours and heat-inactivated at 80°C for 15 min. EBY100, BY4741, and BYDis were transformed according to the FrozenEZ kit instructions (ZymoResearch T2001) and plated on selective media. After 2-3 days, single colonies were picked into selective media to confirm transformation. Prior to induction of Fab expression, yeast were grown up overnight. Confluent cultures were washed in YEP-galactose prior to induction and diluted 1:10 into a final volume of YEP-galactose. EBY100 was induced for 48 h at 20°C with shaking. BYDis expression was tested at a range of temperatures and induction lengths.

CRISPR/Cas-9 mediated genome engineering. Gene-specific gRNAs were selected using GeneiousPrime software tools and cloned into CEN Cas9 plasmids. For

gene deletions, non-coding repair DNA was designed with homology to the 5' and 3' genomic overhangs. OPI1 and the proteases Bar1, Pep4, and Prb1 were knocked out essentially as described previously[125]**.** Deletion was confirmed by genomic PCR.

Fab labeling and selection. Induced cultures were harvested, and 10^6 cells were washed with PBSA buffer (1X PBS, 2 mM EDTA, 0.1% Tween-20, 1% BSA, pH 7.4). Antigen was incubated with cells in 100 μL in PBSA at RT for 1 h, repeated PBSA wash at 4°C, and subsequently labeled with secondary antibodies as appropriate. Cells were washed and resuspended in 300 μL of cold PBSA for flow cytometry on a Sony SA3800. Chaperone libraries were handled as described in section 3.3 and sorted on a Sony SH800.

2.4 DISCUSSION

Here we describe the development of a yeast display strain that efficiently expresses native Fab antibody format to facilitate rapid downstream antibody candidate validation. To achieve this, we implemented a synthetic biology-inspired workflow to allow easy assembly and surface display of immune repertoire libraries. Basing the design of our novel display strain on the common laboratory strain BY4741 made BYDis compatible with modern tools used in yeast biology, and may enable more researchers to adopt yeast surface display as an experimental approach. BYDis is not limited to antibody display, as in theory any protein can be coupled to the AGA2 anchor to achieve surface display. Importantly, BYDis reduces antibody induction time to one-day,

compared to the two-day induction period typically adhered to when using EBY100. This enables faster rounds of selection and quicker identification of candidate antibodies.

We deleted *OPI1* in EBY100 and BYDis and showed that Δ*opi1* might positively influence antibody expression and binding. Further testing using more antibodies of different intrinsic expression abilities in BYDis Δ*opi1* is warranted. Additionally, *OPI1* deletion coupled with expression of chaperones should be explored, as ER enlargement and chaperone upregulation can occur concomitantly in antibody-secreting cells and work cooperatively. Chaperone library selection led to enrichment of a cell population that stained positive for AF-647 in the absence of antigen, suggesting an adhesive phenotype. To avoid aberrant phenotype selection, rationally designed chaperone strains can be constructed and tested instead.

In summary, our discovery platform enables facile high-throughput repertoire interrogation. Our approach is agnostic to the source of immune repertoires, and can equally mine naïve, antigen-experienced, donor-derived, or synthetic libraries. BYDis can yield candidate antibodies in a shorter time frame than previously possible, and the use of Fabs allows more precise characterization and faster downstream development.

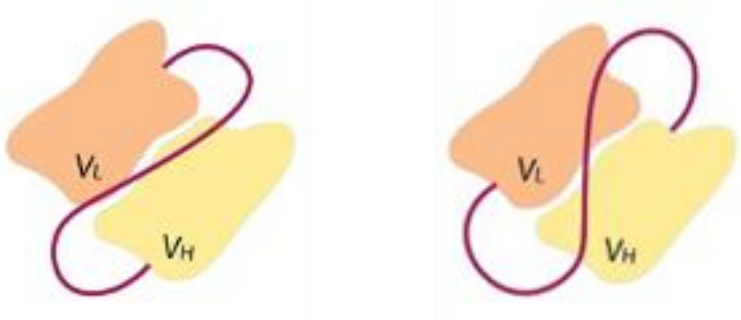

Figure 2.1 ScFv structure. Single chain variable fragments (scFvs) contain only antigen-binding variable domains and a covalent linker. The linker can have multiple configurations.

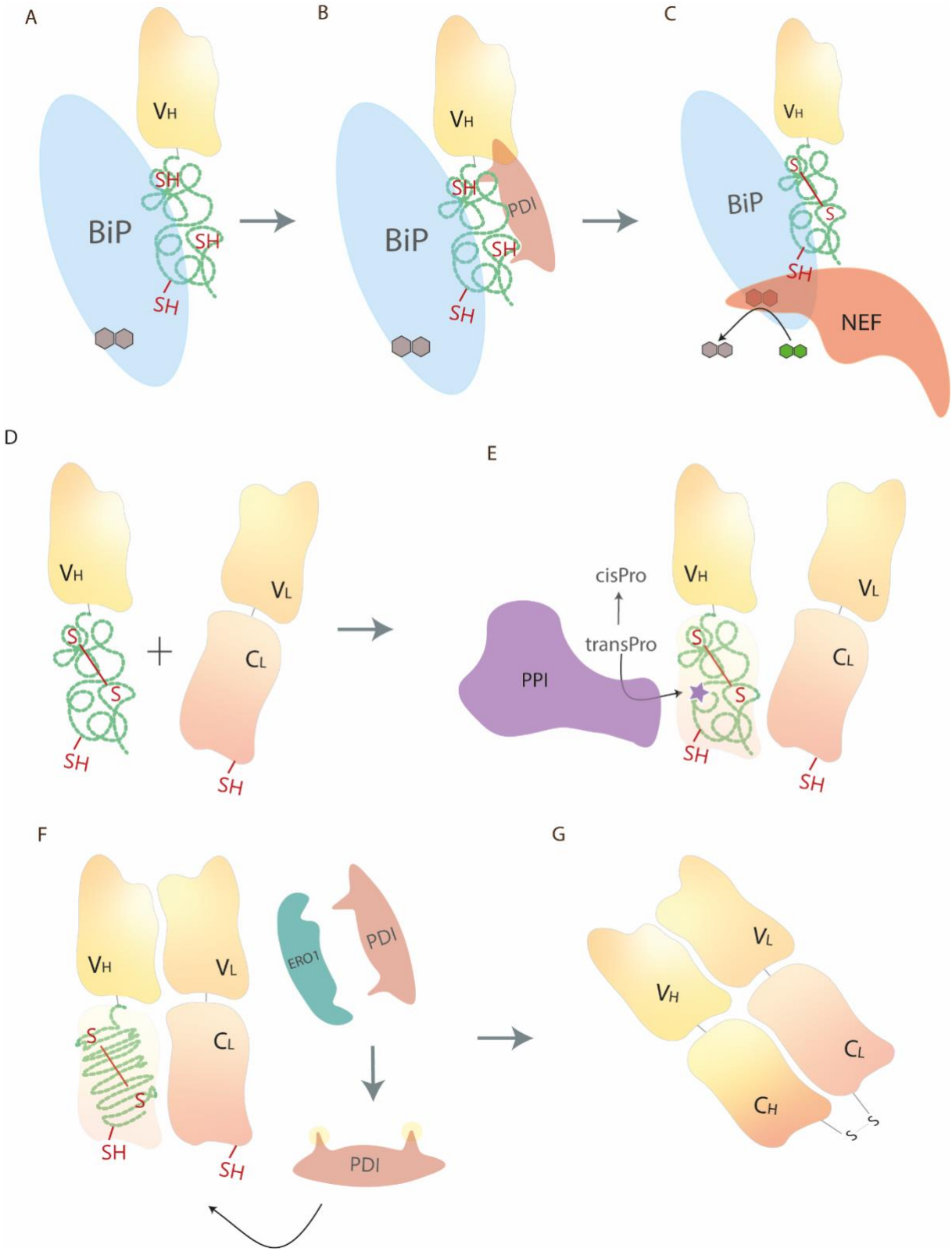

Figure 2.2 Chaperone mediated folding of antibody Fab fragments. A The unfolded

constant domain of the heavy chain (CH1) is bound by the holdase BiP; cysteines are in their reduced state. **B** A PDI chaperone oxidizes the two intradomain disulfide bond-forming cysteines. **C** An NEF cofactor catalyzes the exchange of ADP to ATP in BiP's ATPase domain, causing structural rearrangement to release CH1. **D** The still largely unfolded CH1 domain associates with its native binding partner, the constant region of the light chain (CL). The interaction causes CH1 to begin the collapse into its folded state. **E** A conserved proline residue needs to undergo bond isomerization from the trans to the cis state, assisted by PPIases, to avoid a stable folding intermediate. The CH1 domain collapses into its native immunoglobulin fold. **F** After reoxidation by an oxidoreductase (ERO1, depicted), PDI oxidizes the cysteine residues forming the interdomain disulfide bond. **G** The two subunits then form a stable dimer.

	Function	Chaperone *Homo sapiens*	*S. cerevisiae* homologue	Figure 2.2 Panel	Ref
I	Holdase	BiP	KAR2	A	[126]
II	Protein Disulfide isomerase	PDIA1	PDI1	B	[108]
III	Nucleotide exchange factor	HYOU1 (GRP170)	LHS1	C	[127]
IV	Peptidyl-prolyl Isomerases	CypB	CPR5	E	[128]
V	Oxidoreductase	ERp57	ERO1	F	[129]

Table 2.1 List of chaperones and their function during antibody folding.

All human chaperones listed have been directly linked to antibody folding (references provided), and the yeast chaperones are functional orthologues with endogenous yeast proteins as clients.

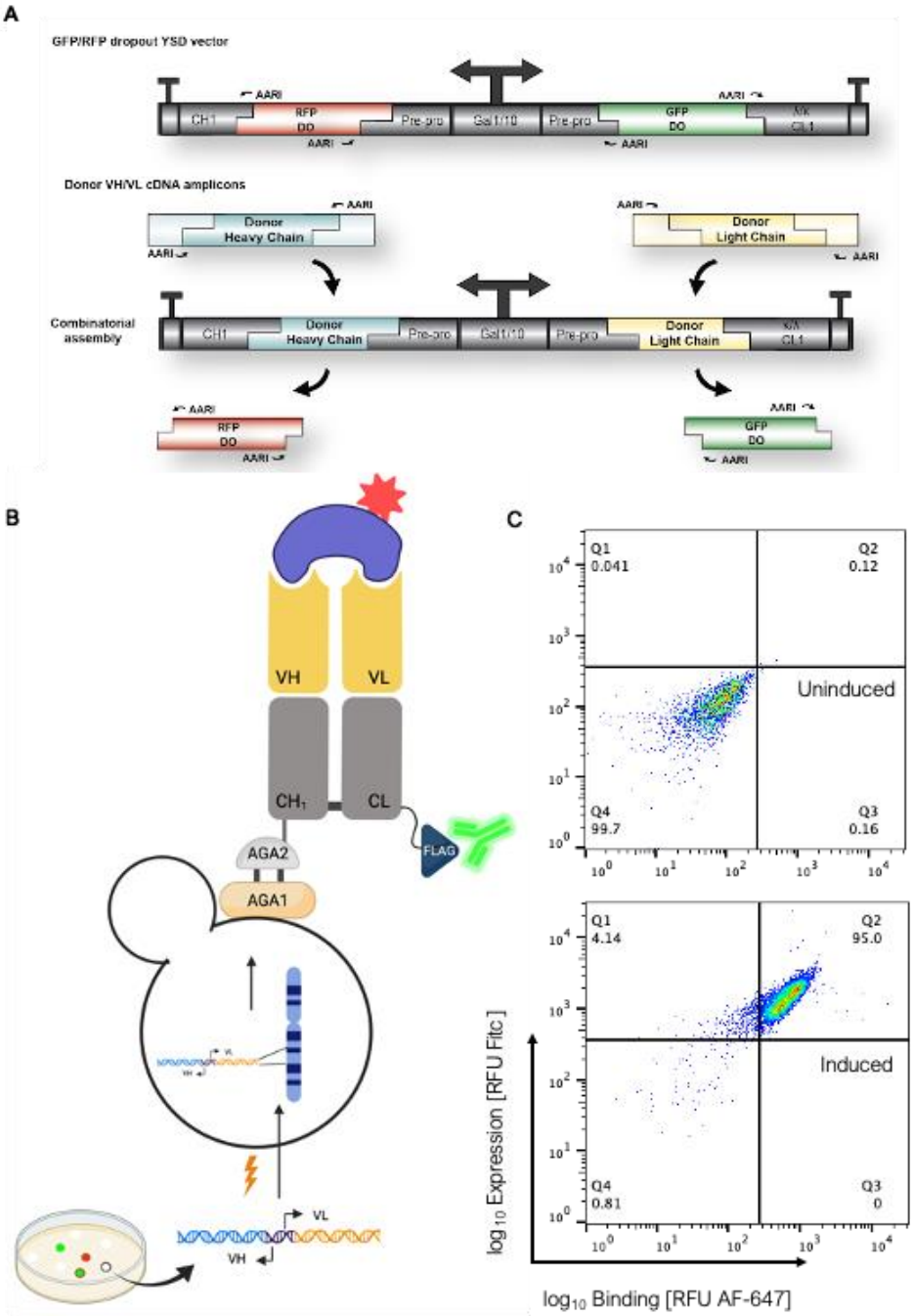

Figure 2.3 Modularly designed antibody display platform. A Antibody display vectors are golden gate compatible and enable green-red screening for correct inserts. **B**, antibody display graphic, created with biorender.com **C** Antibody Fabs integrated at the LEU2 locus display as a homogenous population when induced.

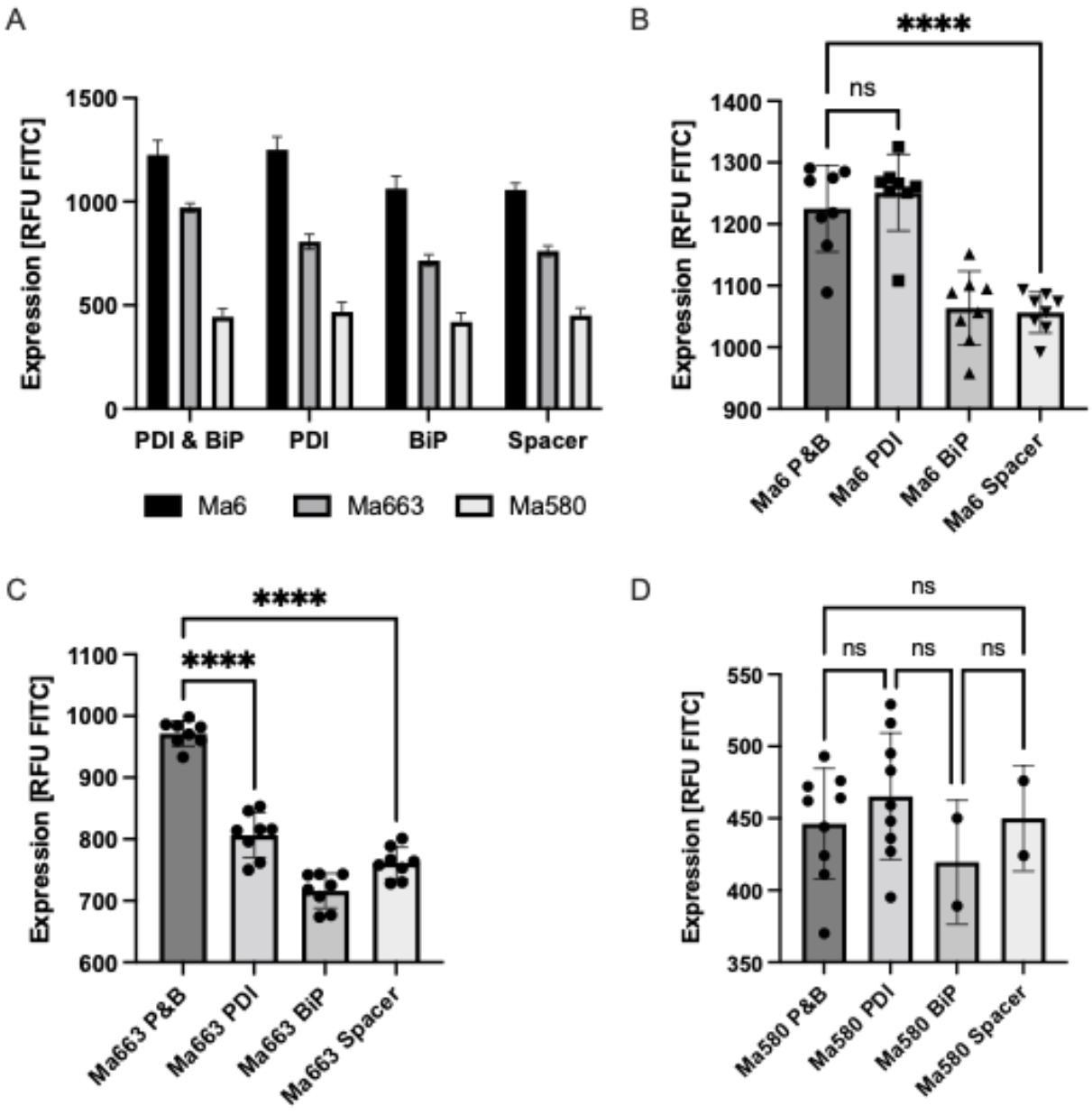

Figure 2.4 Influence of chaperones on antibody Fab expression in EBY100. **A** Expression levels monitored by FITC of Ma6, Ma663, and Ma580 in EBY100 co-expressing chaperones or a non-coding spacer sequence determined by flow cytometry **B** Expression levels of Ma6 are significantly improved in P&B and PDI chaperone strains. **C** Ma663 expression is significantly improved in the double chaperone strain compared to PDI alone. **D** Ma580 expression is not significantly improved by the presence of

chaperones compared to the spacer control. Analyzed through one-way ANOVA with Tukey's multiple comparison (****=$p<0.0001$, ns=not significant).

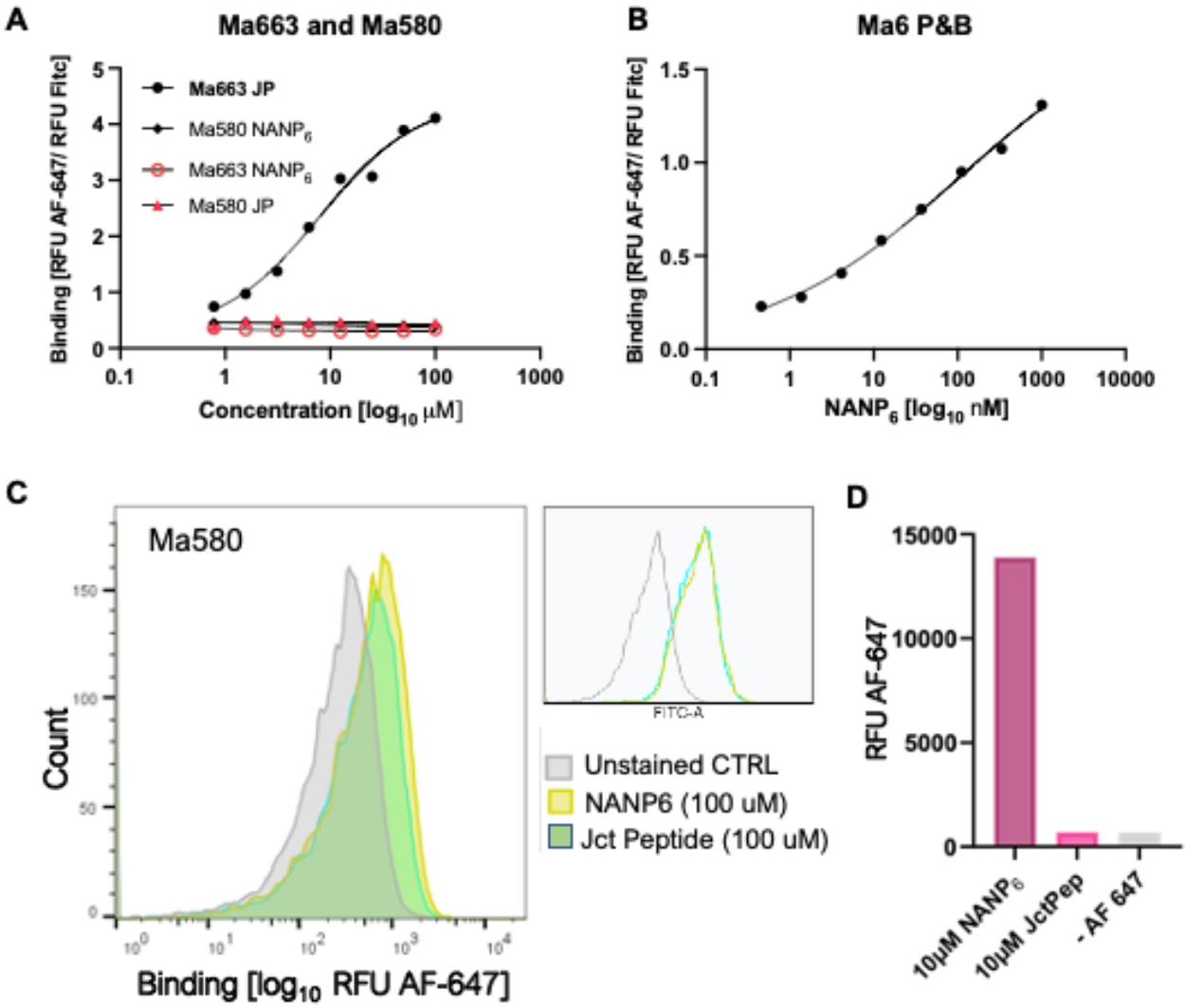

Figure 2.5 Characterization of antibody binding through YSD. **A-D** Yeast-displayed Fab Ma663 binds JP but not NANP6 (A) and no binding to JP or NANP6 can be detected for Ma580 (**A, C**). Ma6 Fab binds NANP6 but not JP (**B, D**).

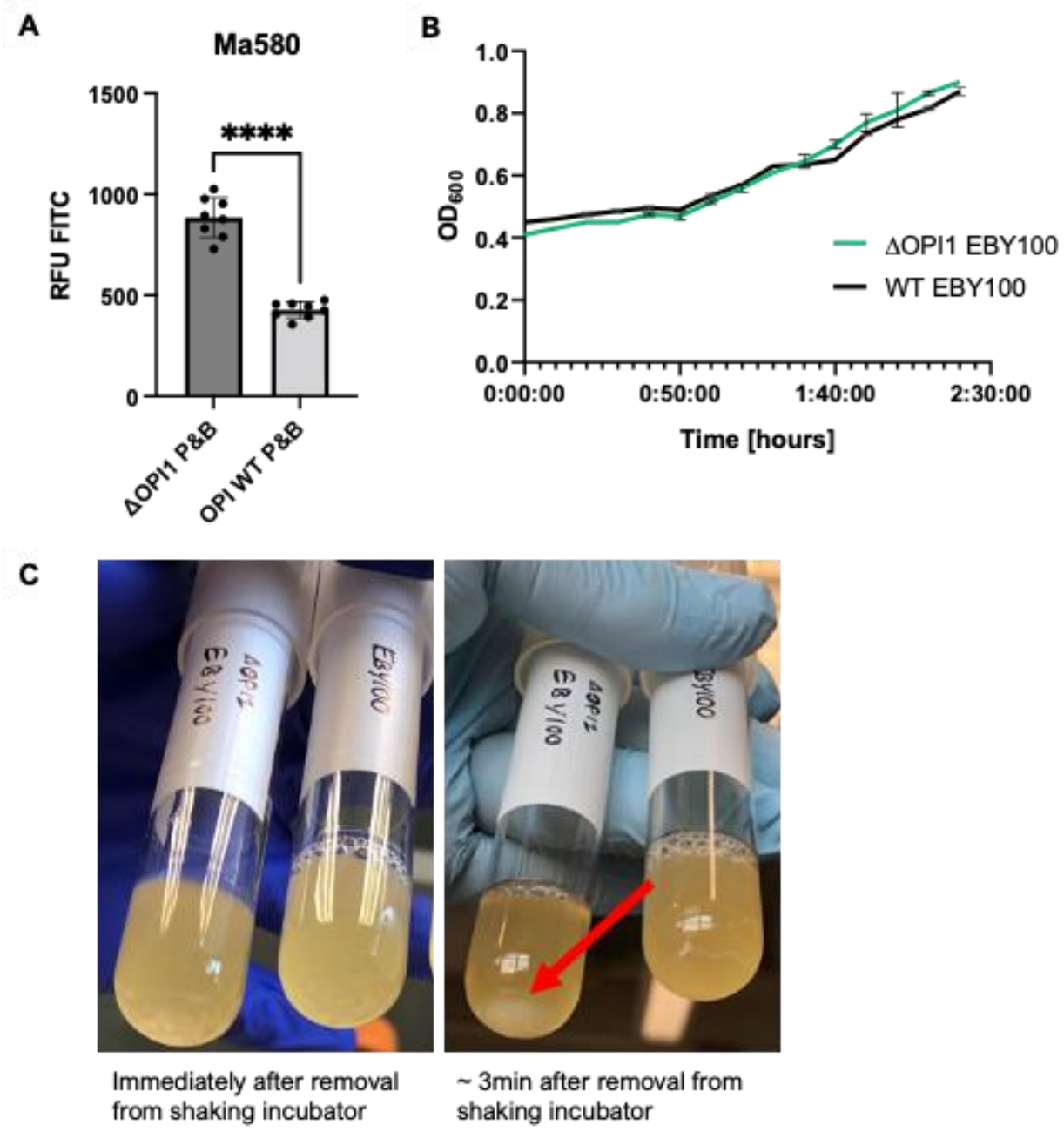

Figure 2.6 Characterization of Δ*opi1* in EBY100. **A** Flow cytometry testing of knockout of OPI1 in EBY100 increases expression of Ma580 in P&B chaperone strain $p<0.0001$ unpaired t-test. **B** Δopi1 shows normal growth behavior compared to WT. **C** However, liquid cultures settle and flocculate at an accelerated rate compared to WT.

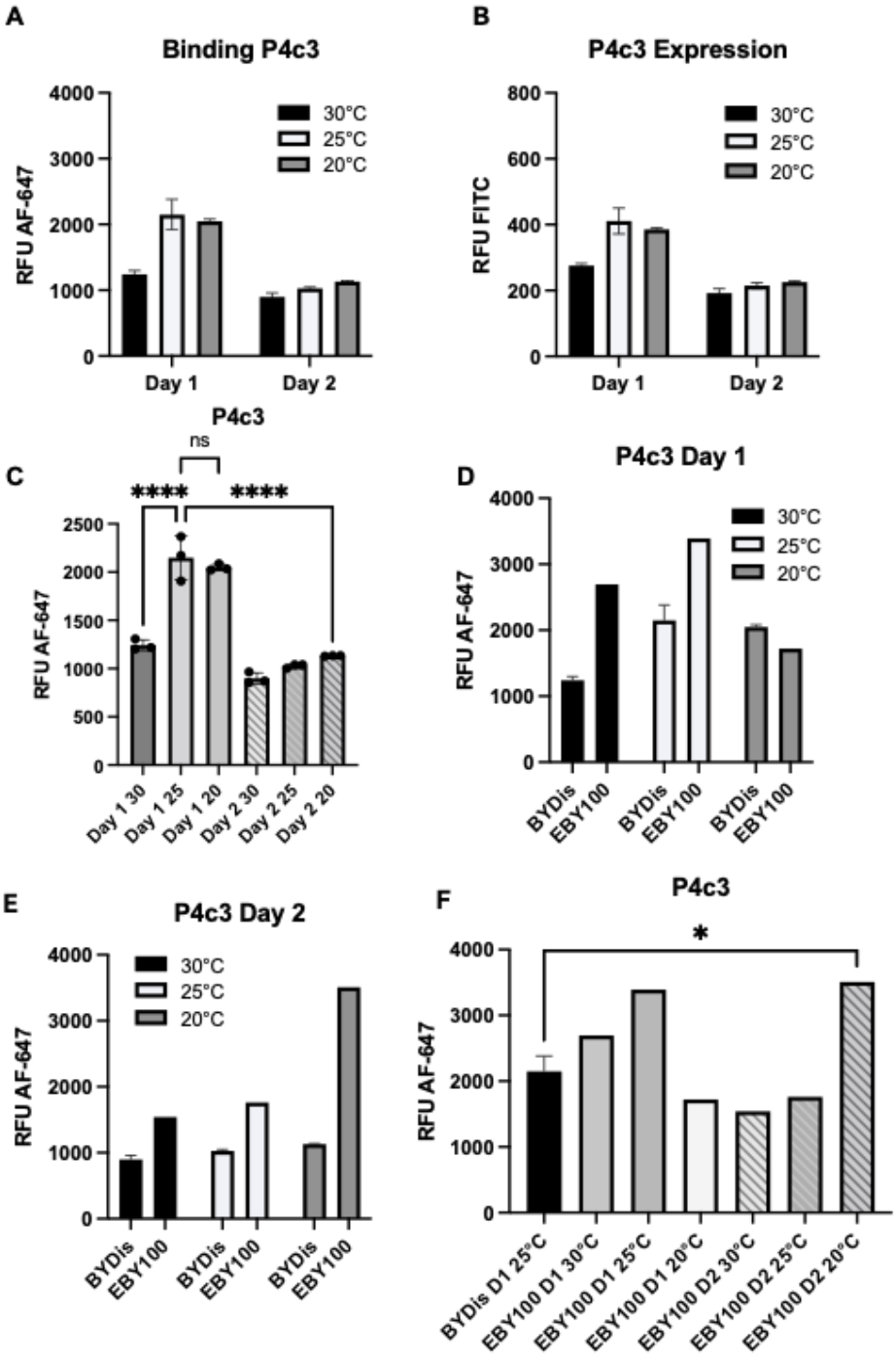

Figure 2.7 Characterization of novel display strain BYDis. **A-C** The Fab P4c3 displayed the best binding and expression when induced at 25°C for one day. Depicted is the mean of biological triplicate repeats, *=p<0.05 ****=p<0.0001 one-way ANOVA,

Tukey's multiple comparison. **D-F** Comparison of BYDis expression and binding to EBY100 WT under equivalent conditions.

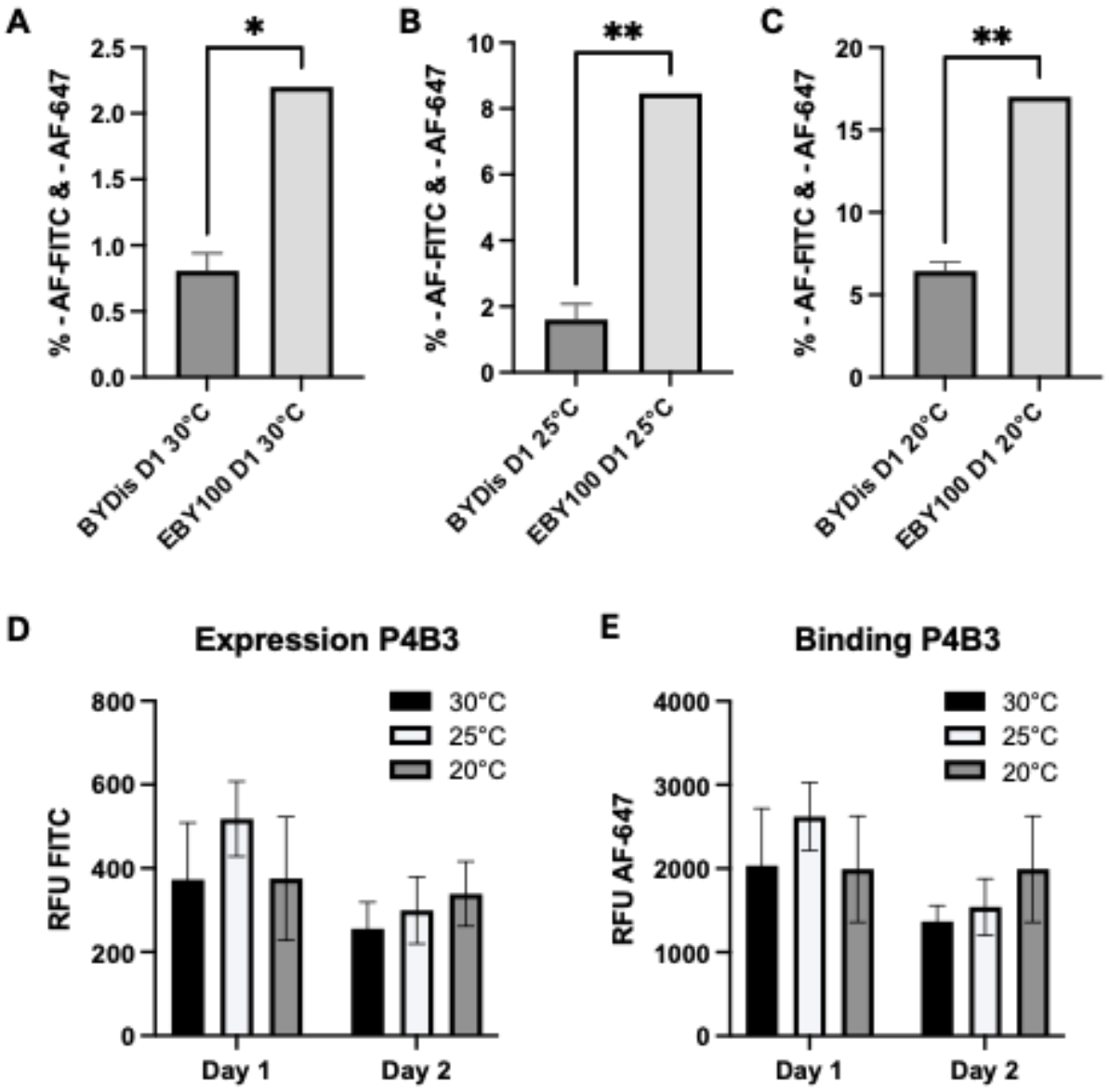

Figure 2.8 BYDis displays a more homogeneous cell population after one-day induction than EBY100. **A-C** Comparison of the percentage of cells in quadrant 4, i.e., non-expressing (-FITC) and non-binding (-AF-647) between BYDis and EBY100 at the different conditions tested. **D-E** Expression and binding profiles of P4b3 in BYDis.

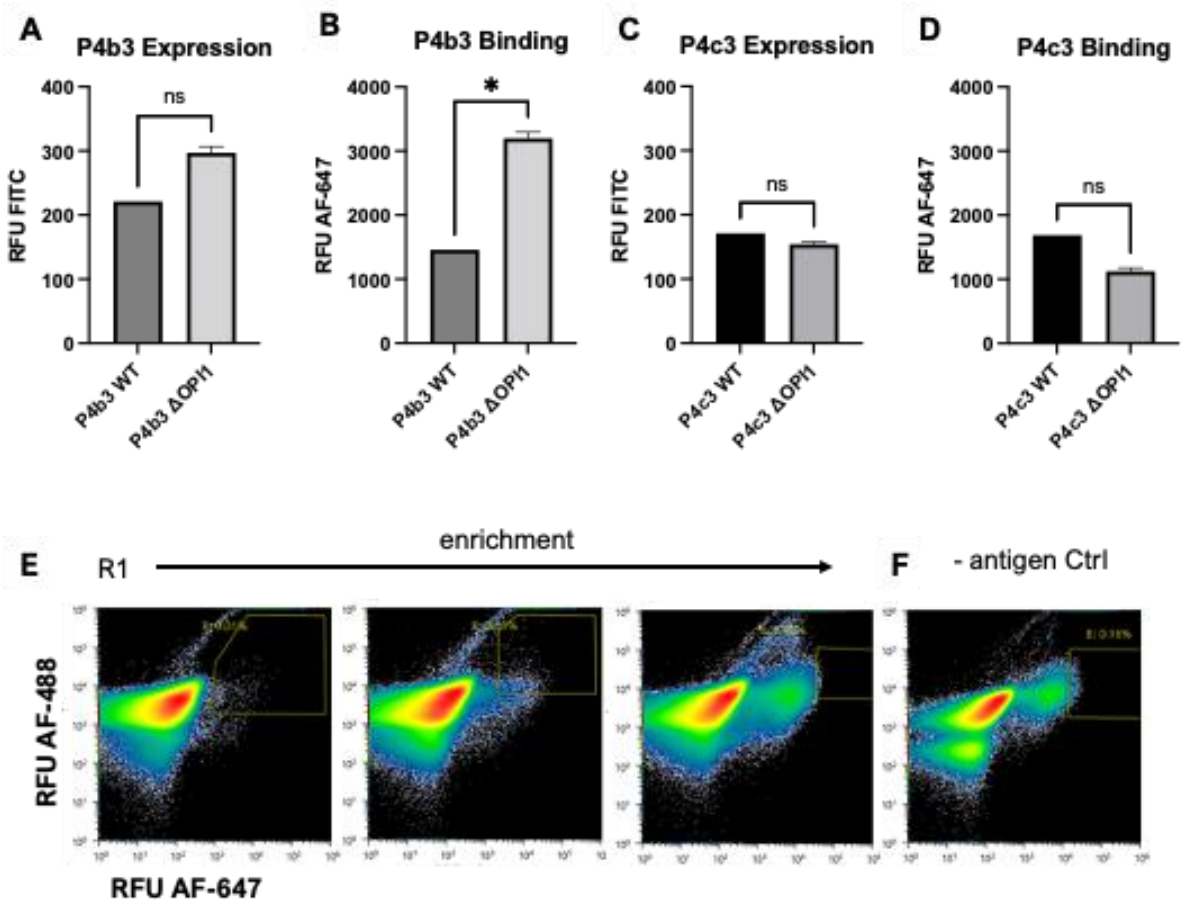

Figure 2.9 Chaperone selection in BYDis Δ*opi1*. **A-D** BYDis Δopi1 improves P4b3 binding but not P4c3,*= $p<0.05$ unpaired t-test. **E** Cell sorting strategy for BYDis Δopi1 transformed with a chaperone library. **F** Increase in AF-488 and AF-647 fluorescence occurs independent of antigen binding.

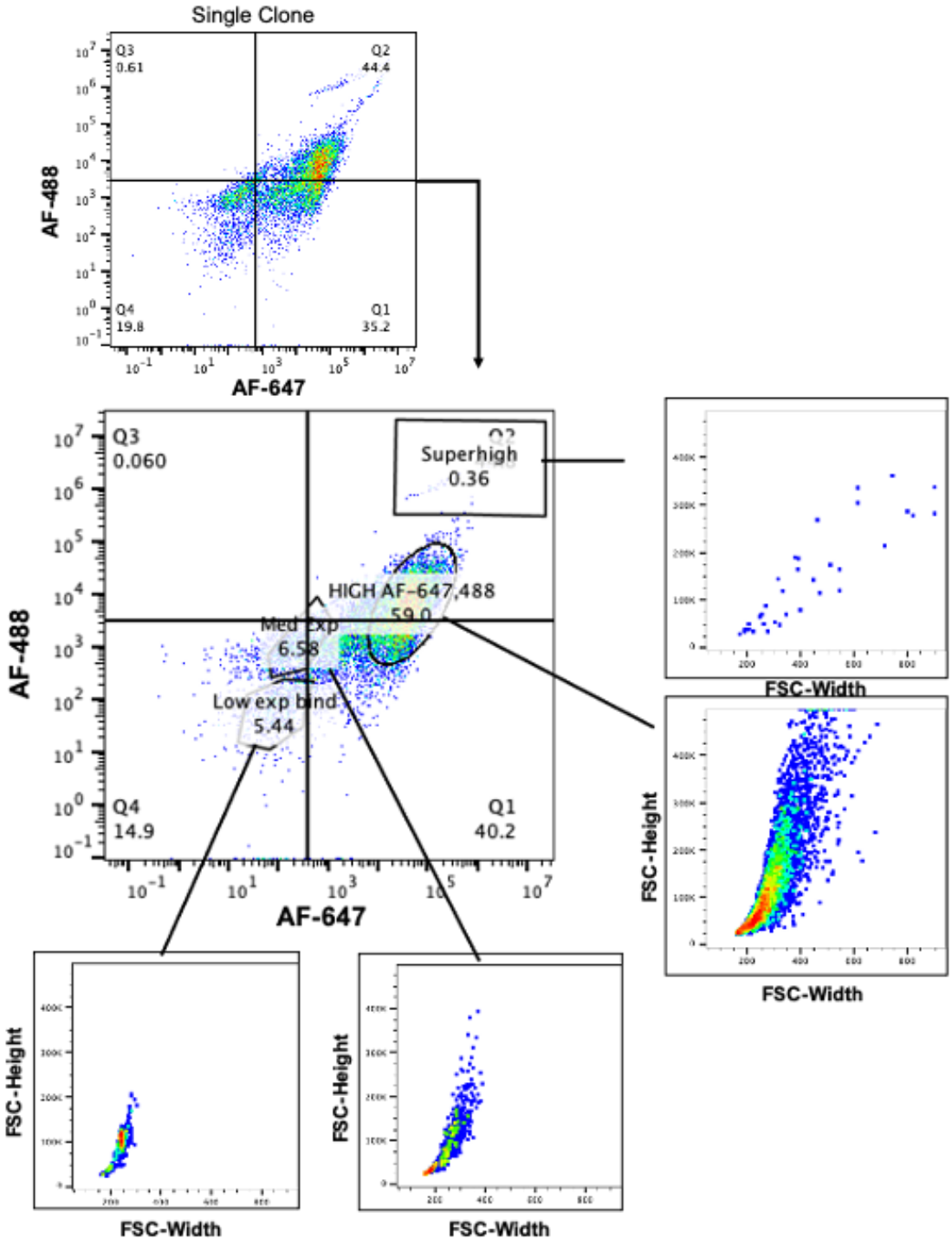

Figure 2.10 AF-488 and AF-647 enriched single clones show heterogeneous cell behavior. Single clones derived from the AF-488 and AF-647 positive population show

discrete cell staining patterns. The distinct cell groups were backgated to determine whether enlarged cell size correlated to greater fluorescent signaling.

CHAPTER 3: SARS-CoV-2 Infection Elicits Convergent Immune Responses and a Neutralizing N-Terminal Domain Antibody Targets an Antigenic Supersite

3.1 INTRODUCTION

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Wuhan, China in late 2019 and its rapid spread across the globe caused a pandemic reminiscent of the 1918 influenza outbreak. The divergence of SARS-CoV-2 from common cold coronavirus strains resulted in the absence of preexisting immunity which, coupled with its airborne nature, enabled accelerated transmission. Though distinguished by its speed of transmission and the severity of symptoms associated with coronavirus disease 19 (COVID-19), SARS-CoV-2 is only the latest in a long line of emerging zoonotic pathogens. Prior to SARS-CoV-2, the World Health Organization warned of the

---

This chapter has been adapted from the manuscript: Jule Goike, Ching-Lin Hsieh, Andrew Horton, Elizabeth C. Gardner, Foteini Bartzoka, Nianshuang Wang, Kamyab Javanmardi, Andrew Herbert, Shawn Abbassi, Rebecca Renberg, Michael J. Johanson, Jose A. Cardona, Thomas Segall-Shapiro, Ling Zhou, Ruth H. Nissly, Abhinay Gontu, Michelle Byrom, Andre C. Maranhao, Anna M. Battenhouse, Varun Gejji, Laura Soto-Sierra, Emma R. Foster, Susan L. Woodard, Zivko L. Nikolov, Jason Lavinder, Will N. Voss, Ankur Annapareddy, Gregory C. Ippolito, Andrew D. Ellington, Edward M. Marcotte, Ilya J. Finkelstein, Randall A. Hughes, James M. Musser, Suresh V. Kuchipudi, Vivek Kapur, George Georgiou, John M. Dye, Daniel R. Boutz, Jason S. McLellan, Jimmy D. Gollihar. Synthetic repertoires derived from convalescent COVID-19 patients enable discovery of SARS-CoV-2 neutralizing antibodies and a novel quaternary binding modality. Biorxiv. doi: https://doi.org/10.1101/2021.04.07.438849

Conceptualization: JG, CH, AH, ECG, DRB, JSM, and JDG; Methodology: JG, CH, AH, ECG, FB, NW, KJ, AH, RR, MJJ, SLW, ZLN, JL, TSS, SVK, VK, RAH, IF, DRB, JSM, and JDG; Investigation: JG, CH, AH, ECG, FB, NW, AA, KJ, AH, WNV, JAC, ACM, RR, MJJ, SLW, ZLN, JL, RAH, SVK, VK, IJF, JMM, DRB, JSM, and JDG; Data Analysis and Interpretation: JG, CH, AH, DRB, NW, AH, JMD, IJF, EMM, JSM, and JDG; Data Curation: JG, CH, AH, ECG, NW, DRB, JSM, AH, JDM, JSM, and JDG; Original Draft: JG, CH, AH, ECG, DRB, JMM, JSM, VK, EMM, and JDG; Review & Editing: JG, CH, AH, ECG, DRB, ADE, EMM, JMM, IJF, GG, JSM, and JDG; Funding: JMD, IJF, GCI, VK, SVK, GG, ADE, JSM and JDG

spread of bird flu (H5N1, avian influenza strain infecting humans)[130], swine flu (2009-2010 outbreak of H1N1)[131], Zika virus (transmission from mosquito to humans, which can cause microcephaly in fetuses, first reported in 2015)[132], Ebola virus (suspected to be transmitted from fruit bats and causes often fatal hemorrhagic fever, most recent outbreak 2014-2016)[133], and the historic 1918 H1N1 influenza strain (suspected avian origin)[134]. Although the zoonotic origin of SARS-CoV-2 is debated, evidence strongly indicates its crossover to humans from bats via an intermediary animal host[135–137]. This recent history of animal-borne diseases spreading to humans emphasizes how vulnerable we are to novel microbial threats, and how urgently we need rapid methods of intervention.

Functional immune repertoire interrogation enables analysis of the primary immune response to emerging pathogens to identify newly elicited antibodies. Protective antibodies can function therapeutically once a patient is infected or be used prophylactically to prevent infection. Antibodies against previous coronavirus strains, such as middle eastern respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV) were largely unable to bind SARS-CoV-2 and the degree of sequence divergence between strains limited cross-reactivity[138]. Coronaviruses have a limited number of surface antigenic targets and only the spike protein (S-protein) elicits neutralizing antibodies[139].

The S-protein is a homo-trimer that initiates viral cell entry by binding angiotensin-converting enzyme 2 (ACE2)[140–142], which is expressed in lung endothelium and other tissues. The S-protein contains S1 and S2 domains[143]. The S2 domain is

anchored in the viral membrane and forms a stalk of alpha-helices before connecting to the S1 domain at its apex. The S2 apex contains a furin cleavage site for post-translational cleavage of S-protein into S1 and S2, and the fusion peptide (FP), which is essential for viral envelop fusion with host membranes. S1 is further functionally divided into the receptor-binding domain (RBD), which engages ACE2 and the N-terminal domain (NTD), which is heavily glycosylated and may contribute to host cell surface adhesion[144,145]. The S-protein is highly dynamic and its ability to switch from a pre-fusion to a post-fusion conformational state is indispensable for SARS-CoV-2 infectivity[146]. The RBD of the trimer can assume compacted or extended RBD 'down' or 'up' conformations. In the up conformation, RBD can bind ACE2 to expose an S2 cleavage site proximal to FP[147]. After S2 cleavage, S-protein can assume its irreversible post-fusion state.

Functional immune repertoire analysis of COVID-19 patients reveals that neutralizing antibodies predominantly target the RBD and NTD of the S1 domain[148–153]. However, selective pressure on these epitopes induces the emergence of SARS-CoV-2 escape mutants. Cell culture experiments demonstrate that the virus can quickly mutate single epitopes to evade individual neutralizing antibodies[154]. Antibody cocktails that combine two or more antibodies targeting different epitopes limit viral escape. A prominent example of a therapeutic mAb cocktail is REGN-COV2, which contains REGN10933 and REGN10987 and targets two distinct RBD epitopes. REGN-COV2 gained FDA emergency use authorization to treat COVID-19[89,92].

Since the spread of the original Wuhan SARS-CoV-2 strain, the virus has displayed a high propensity to evolve fitness-increasing mutations, including D614G in its S2 domain, which arose early in 2020 and quickly outcompeted the Wuhan strain because of its increased infectivity[155]. The glycine substitution reduces spontaneous S1 shedding in the absence of previous receptor docking[156]. While D614G did not affect antibody binding to RBD and NTD epitopes, emerging variants alpha (B.1.1.7) and beta (B.1.351) began to accumulate mutations in these regions. REGN-COV2 continued to neutralize alpha and beta but with reduced potency[157–159]. Similarly, variants circulating in Houston during the first half of 2020 already possessed S-protein mutations that detrimentally affected the binding of neutralizing antibodies elicited to the original Wuhan-Hu-1 S-protein[160]. The trend of increased immune evasion and higher transmissibility continued with the global dissemination of delta in 2021[161]. The aa substitution P681R close to the S-protein's furin cleavage site may increase infectivity by making the site more accessible[162]. However, delta immune evasion is driven by nine amino acid mutations in S-protein that are located throughout the RBD, NTD, and S2[163]. This drift in the antigenic landscape of the spike demonstrates a critical need to identify neutralizing antibodies to a broad spectrum of non-overlapping epitopes on the S-protein to achieve potent and persistent neutralization.

To achieve this goal, we mined immune repertoires of COVID-19 patients using a parallel strategy to discover circulating serum antibodies through immunoglobulin sequencing (Ig-Seq) and yeast surface display-based selections of combinatorially paired

VH:VLs from BCR repertoires. Ig-Seq is a powerful technique to obtain information about the circulating serum antibody repertoire. Once antibodies have been secreted by their B-cells, the genotype to phenotype linkage is lost and sequence determination requires sophisticated proteomic-based methods. Georgiou et al. first developed a workflow that enabled the identification of antigen-specific, circulating antibodies from donor serum[164]. Initially, all serum-secreted antibodies are bound to a Protein A column and antigen-specific immunoglobulins are subsequently separated by affinity chromatography. Protease digestion creates peptide fragments that can be analyzed via mass spectrometry. The mass of each peptide is then correlated to a reference library of VH sequences from peripheral B-cells of the same donor, re-establishing phenotype and genotype linkage. Ig-Seq provides insight into which antibodies actively mediate protection during an infection or respond to immunization[165,166]. Antibodies identified from serum often have desirable therapeutic attributes, such as serum stability.

The YSD arm of our discovery platform displayed combinatorial VH:VL Fab fragments and selected high-affinity binders through multiple rounds of selection. Combined, YSD and Ig-Seq yield a synergistic strategy to discover antibodies from the secreted and circulating repertoire. Especially during primary immune responses, probing both repertoires is advantageous, as optimal sample harvesting time points can only be estimated. Through this strategy, we identified 32 neutralizing antibodies that target a breadth of epitopes, including the RBD, NTD, and elusive S2 subunit. We gained valuable insight into the SARS-CoV-2 antibody response and detected increased usage of

family three heavy V-genes in S-protein-specific antibodies across multiple donors. We report the structural characterization of an NTD supersite-directed antibody and the discovery of an ultrapotent 18 pm IC50 neutralizer.

3.2 RESULTS

3.2.1 **YSD and Ig-Seq Identify SARS-CoV-2-Specific Antibodies from Human Donors**

Our two-pronged approach of Ig-Seq and YSD enabled us to mine the BCR and circulating serum antibody repertoire. Blood samples from symptomatic COVID-19 donors during the first wave of SARS-CoV-2 infections in Austin (TX) were separated into PBMCs and serum. We isolated unpaired heavy and light chain antibody sequences from the PBMCs, and VH and VL sequences were amplified separately by multiplex PCR after reverse transcription of total mRNA. After the addition of restriction enzyme sites 5' and 3' of the variable gene sequences, light and heavy chains were ligated into our YSD vectors and subsequently integrated into *S.cerevisiae* strain EBY100 containing the chaperones BiP and PDI. IgG and IgM amplicon libraries were submitted for MiSeq sequencing.

We tracked the collapse of repertoire diversity and enrichment of SARS-CoV-2 specific antibodies over three rounds of selection with the two proline-stabilized (S2P) ectodomain (ECD) of the Wuhan-Hu-1 S-protein using a diverse gating strategy to account for different expression levels within the Fab library (**Figure 3.1A**). Initially, 62

different IGHV genes were detected in IgG and IgM libraries. Through the selection, non-antigen-specific antibodies were depleted, and different donors showed similar but distinct V-gene enrichment (**Figure 3.1 B**,**C**). Both Donor 1 (D1) and Donor 2 (D2) antigen-specific antibodies were dominated by IGHV3-30 and the very closely related V-genes, IGHV3-30-3 and IGHV3-30-5. This finding is consistent with others noting the increased occurrence of IGHV3-30 antibodies in SARS-CoV-2-specific repertoires compared to healthy donors[167] and specifically in RBD-directed neutralizers[168].

In both D1 and D2, the frequency of IGHV3-30, IGHV3-30-3, and IGHV3-30-5 is higher in IgM antibodies compared to the IgG repertoire. This might indicate that IGHV3-30, -3, and -5 antibodies are newly elicited by SARS-CoV-2 infection, rather than recalled from previous encounters with related coronaviruses, as they have not yet undergone class switching at blood draw. Over the three rounds of selection, IGHV3-30, -3, and -5 accounted for over 40% of mAbs in some samples (**Figure 3.1D**) and the overall IGHV diversity decreased (**Figure 3.1E**). In addition to sequencing each gate, yeast were plated on selective media after sorting to isolate single clones and assess them through phenotyping against the ECD and RBD in a high-throughput flow cytometry assay (**Figure 3.2A**). Candidates for mammalian expression of full-length IgG were selected based on expression-normalized binding (**Figure 3.2B**). IGHV3-30, -3, and -5 were enriched in candidate mAbs from each donor; however, D1's most frequent sequence was IGHV4-34, and D2 additionally showed enrichment for IGHV3-21 (**Figure 3.3A, B**).

At the same time, SARS-CoV-2 S-protein-specific antibodies were isolated from serum through affinity chromatography and subjected to tandem mass spectrometry. We identified 15 distinct CDRH3 peptides for Donor 1 and 21 for Donor 2 in pull-downs against the S-protein ectodomain (ECD) and RBD (**Figure 3.4**). These clonotypes were mapped back to the donors' personalized VH reference libraries, and full-length VH sequences were constructed. Similar to the V-gene distribution found in the antigen-specific YSD-isolated antibodies, Ig-Seq-derived VHs predominately use IGHV3-30, IGHV3-30-3, and IGHV3-30-5 and other IGHV3 family members. Additionally, IGHV1 family members 1-18, 1-24, and 1-69 were observed frequently (**Figure 3.5A**). As light chain partner information is lost during the Ig-Seq workflow, Ig-Seq-derived VH sequences were combinatorially paired with donor VL sequences through ligation into YSD vectors and subjected to selection of VH:VL pairs. VH sequences preferentially paired with IGKV1(D)-39, IGKV4-1, IGLV1-40, and IGLV1-51 (**Figure 3.5B**).

3.2.2 Ig-Seq and YSD Identify Potent Neutralizers Against All SARS-CoV-2 S-Protein Subdomains

We screened candidate mAbs via ELISA against the S2P spike ectodomain (ECD), RBD, NTD, and S2. Binders were selected for live virus neutralization testing by USAMRIID based on their antigen-binding profile. YSD and Ig-Seq-discovered heavy chains showed significant overlap in V-gene usage as expected; however, each technique

identified additional non-overlapping V-gene clonotypes (**Figure 3.6A**), underscoring the usefulness of interrogating the secreted and cellular repertoire for antibody discovery.

Seventeen Ig-Seq-YSD pairs comprising ten clonotypes neutralized authentic SARS-CoV-2. YSD-YSD VH:VL pairing discovered 15 neutralizing VH:VL pairs from 10 clonotypes (**Figure 3.6B**, **Table 3.1**). Of these 32 neutralizing mAbs, the heavy chain V-genes IGHV1-24, and IGHV3-30 comprised the majority of pairs. Additional neutralizing clonotypes identified were IGHV1-69, IGHV3-11, IGHV3-20, IGHV3-23, IGHV3-30-3, IGHV3-66, and IGHV4-31. IGHV1-69, IGHV3-11, IGHV3-20, IGHV3-23, and IGHV3-66 each paired exclusively with one light chain, while IGHV1-24, IGHV3-30, IGHV3-30-3, and IGHV4-31 formed multiple pairs, with light chain V-gene usage varying between clonotypes of the same VH-gene (**Figures 3.7A, B**).

Three distinct IGHV1-24 clonotypes – two derived from Ig-Seq and one from YSD – yielded ten VH:VL pairs, all of which mapped to the NTD. IGHV1-24 VH A7 emerged as a potent neutralizer in combination with four different VLs. Three of the partnered light chains were highly similar variations of IGLV1-51 (VH:VL pairs A7V3, A7V131, A7V132) and the resulting mAbs ranged in IC50 from 0.305 nM to 1.39 nM. The most potent A7 pair used the less frequently observed IGLV2-8 with an IC50 of 0.018 nM. Notably, A7 exclusively paired with lambda light chains. A different ultrapotent IGHV1-24 VH, 12C, also formed a highly neutralizing pair with IGLV1-51 (12C114, IC50=0.35 nM) but was able to pair with more diverse VL genes, including two pairs with the Kappa V-gene IGKV1-39 (12C130 and 12C113). Interestingly, 12C also

formed two pairs with IGLV1-40, 12C115, and 12C112, which varied significantly in IC50 from 1.929 nM to over 300 nM, despite largely identical sequences. This finding is inconsistent with the tightly spaced IC50 distribution of the IGLV1-51-utilizing A7 neutralizing pairs. The YSD-derived IGHV1-24 VH Q7 also paired with a lambda V-gene, IGLV3-19, and displayed an IC50 of 0.98 nM, making it the sixth sub-nanomolar IGHV1-24-based neutralizer.

Of the ten IGHV3-30 based neutralizers, four clonotypes were identified via YSD and five through Ig-Seq. The nine IGHV3-30 CDRH3 sequences possessed remarkable diversity and ranged in length from 6 aa (9C) to an uncommonly long 27 aa (8C). While all IGHV1-24 VH-based neutralizers targeted the NTD, nine IGHV3-30 VHs bound different subdomains of S-protein, including the RBD, NTD, and S2. Despite the diversity in CDRH3 and epitope specificity, the ten IGHV3-30 neutralizers showed a preference for a narrow range of light chain V-genes, specifically IGLV1-40 and IGKV1-39, which they paired with eight times. IGHV3-30 pairs displayed weaker neutralization potency compared to IGHV1-24 mAbs, varying from 20 nM to >300 nM IC50s.

3.2.3 A7V3 Binds NTD and Protects SARS-CoV-2 Challenged Mice

We identified A7V3 as a promising candidate early in the screening process and observed A7's V-gene IGHV1-24 in combination with an IGLV1-51-utilizing light chain several times during selection. Three additional neutralizers also possess this VH:VL pairing. Despite the strong preference for IGLV1-51, A7V3 is likely a non-cognate VH:VL pair, as its Ig-Seq-derived VH A7 could be traced to Donor 1 and its IGLV1-51

light chain to Donor 2. To determine how A7V3 mediates binding to the S-protein and investigate the potential significance of an IGLV1-51 pairing, our collaborators solved a cryo-EM structure of A7V3 Fabs bound to the SARS-CoV-2 S-protein.

The 3.0 Å-resolution map confirmed ELISA-based epitope mapping and showed one A7V3 Fab fragment bound to each NTD of the spike trimer (**Figure 3.8A**). Structure refinement efforts focused on one Fab with clear density and produced a defined binding interface with the NTD (**Figure 3.8B**). Examination of the binding interaction revealed an analogous role of A7V3's heavy chain in N3 and N5 loop binding, as described for the first structurally characterized NTD-directed mAb 4A8[152]. 4A8 also possessed an IGHV1-24 heavy chain but had a longer CDRH3 and paired with IGKV2-24. A7V3's CDRH1, CDRH3, and CDRL2 formed a clamp-like conformation that pinned the highly flexible NTD loop 5 (residues 246-260) into a fixed position (**Figure 3.8C**). Tyr49 and Pro55 in CDRL2 cooperated with CDRH3 Trp100d to encase Pro251 at the tip of the N5 loop through hydrophobic interactions. Hydrogen bonds contributed to binding through CDRH3's Asp101 interaction with Tyr248's hydroxyl group. Glu31 in the CDRH2 likely forms a salt bridge with Arg246, consistent with interactions reported for 4A8[152]. Pro97 and Phe98 of the CDRH3 were inserted into a ridge formed between the N5 and N3 loop (**Figure 3.8D**). Contacts with the N3 loop (residues 141-156) were less extensive and mediated by the CDRH1 and CDRH2, both in A7V3 and 4A8. Overlay of the CDRHs of A7V3 and 4A8 demonstrate how similar binding interactions were in the common sequence space (CDRH1 and CDRH2) before diverting because of the distinct CDRH3

architectures (**Figure 3.8E**). For both A7V3 and 4A8, the light chain made limited contacts with the NTD and likely contributed more significantly through structural stability for the VH.

Collaborators at Professor Reizis's laboratory at the New York University School of Medicine performed protection assays in mice. Mice were challenged with SARS-CoV-2 in the absence or presence of A7V3 and their vitals and disease progression were assessed. On day four after infection, mice receiving A7V3 had lower lung virus titers (**Figure 3.9A**) and showed reduced levels of weight loss compared to controls (**Figure 3.9B**). Most importantly, mice receiving A7V3 maintained a 100% probability of survival over four days, whereas unprotected mice dropped to 50% (**Figure 3.9C**), indicating that A7V3, an NTD-directed antibody, protects mice from disease.

3.3 Materials and Methods

Strains and media. *E. coli* strain DH10B was used for all routine cloning of yeast and mammalian constructs. Yeast strain EBY100 (MATa AGA1::GAL1-AGA1::URA3 ura3-52 trp1 leu2-delta200 his3-delta200 pep4::HIS3 prbd1.6R can1 GAL) was acquired from ATCC (cat. no. MYA-4941) and used for antibody expression and selection. To improve antibody expression, the human chaperones BiP (binding immunoglobulin protein) and PDI (protein disulfide isomerase) were genomically integrated as an expression cassette in the HO locus. Yeast were grown in rich medium (YPD; Takara, cat. no. 630409) or in selective medium for leucine prototrophs after library transformation (Takara cat. no. 630310). YEP-galactose was used for expression of

displayed antibody libraries (1% yeast extract, 1% bacto-peptone, 0.5% NaCl, 2% galactose, 0.2% glucose).

Antigens and antibodies. Stabilized spike ECD antigen S-2P was biotinylated using the EZ-link kit (Thermo Scientific, cat. no. 21435) and labeled with streptavidin-AF647 (Invitrogen, cat. no. S32357). RBD was labeled with a mouse anti-human Fc-AF647 (Southern Biotech, cat. no. 9042-31). Fab library light chains were labeled with anti-FLAG M2-FITC (Sigma, cat. no. F4049).

Donors. Blood was collected from three PCR-confirmed, symptomatic patients. Donors 1 and 2 samples were collected at day 12 post symptom onset. None of the donors were hospitalized or experienced severe disease. PBMCs and plasma were both collected by density gradient centrifugation using Histopaque-1077 (Sigma-Aldrich).

Preparation of serum antibodies for Ig-Seq proteomic analysis. Serum samples were prepared for Ig-Seq analysis as previously described[169]. Briefly, total IgG was isolated from plasma using Pierce Protein G Plus Agarose (Pierce Thermo Fisher Scientific) and cleaved into F(ab')2 fragments with IdeS protease. Antigen-specific F(ab')2 was enriched by affinity chromatography against recombinant SARS-CoV-2 S-2P or RBD protein cross-linked to NHS-activated agarose resin (Thermo Fisher Scientific). Eluted F(ab')2 fractions were concentrated by vacuum centrifugation and prepared for mass spectrometry-based proteomic analysis as previously described[165].

LC-MS/MS analysis of antigen-enriched antibodies. Liquid chromatography-tandem mass spectrometry analysis was carried out on a Dionex Ultimate 3000

RSLCnano system coupled to an Orbitrap Fusion Lumos Mass Spectrometer (Thermo Scientific). Samples were loaded onto an Acclaim PepMap 100 trap column (75 μm × 2 cm; Thermo Scientific) and separated on an Acclaim PepMap RSLC C18 column (75 μm × 25 cm; Thermo Scientific) with a 3%-40% acetonitrile gradient over 60 min at a flow-rate of 300 nL/min. Peptides were eluted directly into the Lumos mass spectrometer using a nano-electrospray source. Mass spectra were acquired in data-dependent mode with a 3 s-cycle time. Full (MS1) scans were collected by FTMS at 120,000 resolution (375-1600 m/z, AGC target = 5E5). Parent ions with a positive charge state of 2-6 and minimum intensity of 3.4E4 were isolated by quadrupole (1 m/z isolation window) and fragmented by HCD (stepped collision energy = 30+/-3%). Fragmentation (MS2) scans collected by ITMS (rapid scan rate, AGC target = 1E4). Selected ions and related isotopes were dynamically excluded for 20 s (mass tolerance = +/-10ppm).

Antibody variable chain sequencing. Peripheral blood mononuclear cells from processed donor samples were provided in Trizol as a kind gift from Dr. Gregory C. Ippolito[153]. IgG and IgM VH and VL cDNA libraries were separately amplified from PBMC RNA of three donors and sequenced to create donor-specific reference databases, from which the complete amino acid sequences of serum IgG proteins could subsequently be determined based on their mass spectral identifications

Ig-Seq MS data analysis. Mass spectra were analyzed using Proteome Discoverer 2.2 software (Thermo Scientific). Precursor masses were first recalibrated with the Spectrum File RC node using a consensus human reference proteome database

(UniProt) with common contaminants (MaxQuant) and precursor mass tolerance of 20 ppm. Recalibrated mass spectra were searched against a custom database for each donor consisting of donor-derived VH sequences, VL sequences, and the human and contaminant sequences using the Sequest HT node. Mass tolerances of 5 ppm (precursor) and 0.6 Da (fragment) were used. Static carbamidomethylation of cysteine (+57.021 Da) and dynamic oxidation of methionine (+15.995 Da) were considered. False discovery rates for peptide-spectrum matches (PSMs) were estimated by decoy-based error modeling through the Percolator node. Label-free quantitation (LFQ) abundances were calculated from precursor areas using the Minora Feature Detector and Precursor Ions Quantifier nodes.

Resulting PSMs were filtered as previously described[165]. Briefly, peptide sequences differing only by isoleucine/leucine substitution were considered equivalent and combined into a single PSM. PSMs were re-ranked by posterior error probability, q-value, and Xscore. Only top-ranked, high-confidence PSMs (FDR <1%) were retained for each scan. If two or more PSMs had identical top-ranked scores, they were considered ambiguous and removed. PSMs for the same peptide sequence were summed and the average mass deviation (AMD) was calculated for each peptide. Peptides with AMD greater than 2 ppm were filtered out. Peptides mapping to VH sequences from a single clono-group were considered clono-specific. Clono-specific peptides overlapping the CDR3 sequence by four amino acids or more were considered CDR3-informative.

For each clono-group, PSMs and LFQ abundances of clono-specific CDR3-informative peptides were summed. Ratios of elution:flow-through PSMs and LFQ abundances were calculated; only clono-groups with both ratios >5 were considered elution-specific.

Library assembly and bacterial transformation. Donor B-cell VH and VL amplicons were amplified via PCR to include adapters for cloning into yeast expression vectors. Assembly into the yeast kappa and lambda expression vectors was done via Golden Gate assembly. Library assemblies were prepared in 20 µL reactions as follows: 2 µL 10X AARI buffer (Thermo Fisher Scientific, cat. no. B27), 0.4 µL 50X oligo buffer (ThermoFisher Scientific, cat. no. ER1582), 0.2 µL 100 mM ATP (ThermoFisher Scientific, cat. no. R0441), 20 fmol backbone DNA, 40 fmol VH and VL amplicons, 0.5 µL (2 U/µL) AARI endonuclease (ThermoFisher Scientific, cat. no. ER1582), and 0.5 µL T7 ligase (3000 U/µL) (NEB, cat. no. M0318). Each assembly was scaled up to 16 total reactions in 8-well strips. Thermocycling consisted of the following protocol: 37°C, 15 min; 37°C, 2 min, 16°C, 1 min; go to step 2, x74; 37°C, 60 min; 80°C, 15 min; hold at 4°C. Assemblies were consolidated and column purified using Promega Binding Solution (Promega, cat. no. A9303) to bind DNA to a Zymo-spin II column (Zymo Research, cat. no. C1008). The column was washed twice with DNA Wash buffer (Zymo Research, cat. no. D4003) and eluted in 30 µL nuclease-free water. For library transformations, DH10B *E. coli* cells were diluted 1:100 from confluent culture into 50 mL Superior broth (AthenaES, cat. no. 0105). When cells reached an OD600 of 0.4-0.6, they were washed

3X with cold 10% glycerol and resuspended to a final volume of 600 µL. The purified library was added to cells and electroporated at 1.8 kV in an *E. coli* Pulser electroporator (Bio-Rad) using Genepulser 0.2 cm-cuvettes (Bio-Rad, cat. no. 1652086) at 200 µL per transformation.

Library transformation into yeast and protein expression. Purified libraries were linearized for integration into the yeast genome via homologous recombination at the Leu2 locus. For each 1 µg library plasmid, 0.5 µL NotI (10 units/µL) (NEB, cat. no. R0189) was used with the supplied buffer 3.1 in 10 µL. Reactions were incubated at 37°C overnight and heat inactivated at 80°C for 20 min. Digests were pooled and column purified as described in previous sections and eluted in 25 µL nuclease-free water. Our strain was electroporated as described elsewhere[170]. We found that 10 µg of linearized DNA was sufficient for library sizes of 106, and that library sizes could reach >107 with 20 µg DNA. Transformed yeast were recovered in SD-Leu medium (see "strains and media" section). Libraries were passaged once at 1:100 before protein expression to reduce contamination from untransformed cells. To express Fab libraries, yeast were washed in YEP-galactose (see "strains and media") and diluted 1:10 into 10 mL final volume. Cells were induced for 48 h at 20°C with shaking.

Fab library labeling and selection. Expressed yeast libraries were harvested at 100 µL (representing approximately 107 cells) and washed with PBSA buffer (1X PBS, 2 mM EDTA, 0.1% Tween-20, 1% BSA, pH 7.4). Antigen was incubated with cells in 1 mL in PBSA at 200 nM at RT for 1 h, washed with PBSA at 4°C, and labeled with

secondary antibodies (mouse anti-human FITC, 1:100; streptavidin-AF647, 1:100; mouse anti-human Fc-AF647, 1:50). Cells were washed 2X and resuspended in 2 mL in cold PBSA for sorting. Cell sorting was performed using a Sony SH800 fluorescent cell sorter. For first round libraries, 107 events were sorted into 2 mL SD-Leu medium supplemented with penicillin/streptomycin (Gibco, cat. no. 15140122). Cells were recovered by shaking incubation for 1-2 days for further rounds of selection or plated directly for phenotyping clones.

Next-generation sequencing. Genome extraction was performed on yeast cultures of libraries and sorted rounds underwent genome extraction using a commercial kit (Promega, cat. no. A1120) with zymolyase (Zymo Research, cat. no. E1004). 100 ng genomic template was used to amplify the heavy and light chains separately or as one amplicon for short or long-read sequencing, respectively. For amplification of heavy chain genes only, primers JG.VHVLK.F and JG.VH.R were used. For amplification of light chain genes only, primers JG.VL.F and JG.VHVK.R or JG.VHVL.R were used for kappa and lambda vectors, respectively. For amplification of paired genes, the primers JG.VHVLK.F and JG.VHVK.R or JG.VHVL.R were used. Amplicons were column purified and deep sequenced with an iSeq. In parallel, we obtained ~1.8 kb sequences spanning the entire VH and VL using MinION nanopore sequencing (Oxford Nanopore Technologies Ltd., MinION R10.3).

Colony PCR and Sanger sequencing. Sorted yeast populations were plated on SD -Leu and 8-32 colonies per plate were picked into 2 mL-microplates either by hand or

using a QPIX 420 (Molecular Devices) automatic colony picker. Cultures were grown at 1000 rpm at 3-mm orbit at 30°C overnight. Cells (20 μL) were transferred to a fresh microplate and washed with 1 mL TE buffer (10 mM Tris, 1 mM EDTA). Cells were incubated with 20 μL zymolyase solution (5 mg/mL zymolyase, 100T in TE) at 37°C for 1 h. Cells (5 μL) were then used in colony PCR to amplify the paired heavy and light chains. Amplicons were column purified with the Wizard SV 96 PCR Clean-Up System (Promega, cat. no. A9342), and yields were quantified with a Nanodrop spectrophotometer or the Quant-it Broad-Range dsDNA kit (Invitrogen, cat. no. Q33130). Approximately 10 ng (2.5-5 μL) of purified PCR products were then subjected to Sanger sequencing.

Long-read sequencing. Sequencing libraries were prepared from 18 amplicon samples using the Native Barcoding Kit (Oxford Nanopore Technologies; cat. no. EXP-NBD103) paired with the Ligation Sequencing Kit (Oxford Nanopore Technologies; cat. no. SQK-LSK109) according to the manufacturer's directions. Between four and eight sequencing libraries per flow cell were pooled for sequencing on three MinION flow cells (Oxford Nanopore Technologies; R9.4.1) for 72 h on an Oxford Nanopore Technologies MinION Mk1B device (Oxford Nanopore Technologies). Raw data was basecalled using the high accuracy model in Guppy (v.3.5.2).

Short-read sequencing (phenotyping plates). Sequencing libraries were prepared from 308 amplicon samples using the Nextera DNA Flex Library Preparation kit (Illumina; cat. no. 20018705) according to the manufacturer's directions. Sequencing

libraries were pooled and sequenced (2x151 bp) on an iSeq 100 (Illumina; California, USA) using iSeq 100 i1 Reagents v.1 (Illumina; cat. no. 20021533).

Cryo-EM sample preparation and data collection. Purified SARS-CoV-2 S with three RBDs covalently trapped in the down conformations (HexaPro-RBD-down variant, S383C/D985C)(16–18) at 0.2 mg/mL, complexed with a 2-fold molar excess of Fab A7V3, was deposited on plasma-cleaned UltrAuFoil 1.2/1.3 grids before being blotted for 3 s with -4 force in a Vitrobot Mark IV and plunge-frozen into liquid ethane. For the HexaPro-RBD-down-A7V3 sample, 3,636 micrographs were collected from a single grid. FEI Titan Krios equipped with a K3 direct electron detector (Gatan) was used for imaging. Data were collected at a magnification of 22,500x, corresponding to a calibrated pixel size of 1.07 Å/pix.

Cryo-EM data processing. Gain reference- and motion-corrected micrographs processed Warp[171] were imported into cryoSPARC v2.15.0[172], which was used to perform CTF correction, micrograph curation, particle picking, and particle curation via iterative rounds of 2D classification. The final global reconstructions were then obtained via *ab initio* reconstruction, iterative rounds of heterogeneous refinement, and subsequently non-uniform homogeneous refinement of final classes with C1 symmetry. For the HexaPro-RBD-down-A7V3 sample, C3 symmetry was attempted in the initial refinement process. Given the low occupancy of the Fabs on a trimeric spike, C1 symmetry was used for the final runs of heterogeneous and homogeneous refinement. To better resolve the Fab-spike interfaces, both datasets were subjected to particle

subtraction and focused refinement. Finally, both global and focused maps were sharpened using DeepEMhancer[173]. For A7V3-NTD model building, we used an NTD from the 4A8 complexed spike structure (PDB ID: 7C2L[174]) and a homologous Fab structure (PDB ID: 6IEK) as an initial model to build into map density using UCSF ChimeraX[175]. Both models were further built and iteratively refined using a combination of Coot[176], Phenix[177], and ISOLDE[178].

Antibody protection assay in MA10 mouse model. The MA10 mouse model was constructed previously as a mouse ACE2 adapted pathogenic SARS-CoV-2 strain for BALB/c mice[179,180]. Twelve-month-old female BALB/c mice (n=5) received 200 µg/mouse A7V3 or PBS control through intraperitoneal injection. Twelve hours post-injection, mice received lethal doses (10^5 or 10^4 PFU) of MA10 intranasally. Their body weight was monitored daily until euthanasia on day 4 post-infection by isoflurane overdose. The right caudal lung lobe was excised and preserved in PBS at -80°C. Viral lung titers were determined through plaque assays performed on Vero E6 cells. The UNC Chapel Hill IACUC protocol for mouse research is 20-114.

3.4 DISCUSSION

We used an integrative approach to discover 32 neutralizing antibodies with specificity for each S-protein subdomain. We report a repertoire polarization towards heavy chains of the IGHV3 family, specifically IGHV3-30, in two SARS-CoV-2-infected donors and the discovery of IGHV1-24 heavy chains with potent NTD-directed neutralizing activity. These findings are consistent with other SARS-CoV-2 antibody

discovery campaigns that noted increases in IGHV3 usage and described IGHV1-24 NTD binders (e.g., 4A8), indicating that SARS-CoV-2 evokes comparable immune responses even in ethnically diverse patients. We describe the binding mode of A7V3 to the N3 and N5 peptide loops of the NTD, which protects mice from developing COVID-19 upon SARS-CoV-2 challenge. Reports of neutralization and protection through NTD-directed antibodies were somewhat surprising early in the pandemic, as our functional understanding of the NTD is incomplete and its role in infection is unknown. Notwithstanding, the N3 and N5 loops have been identified as antigenic supersites that consistently evoke convergent immune responses across different patients[151,181]. In contrast to the rest of the NTD, its loops are not heavily glycosylated, which may explain the focus of antibody recognition at this site. NTD-directed antibodies provide neutralization and infection protection consistent with RBD-targeting antibodies. Up to 20% of donor serum antibodies show activity towards the NTD[151]. Unfortunately, early emerging variants, including alpha, beta, and gamma, acquired mutations in the N-loops, causing loop architecture restructuring, thereby circumventing protection through previously elicited NTD-directed antibodies[182].

When omicron emerged in November of 2021, it had accumulated an unprecedented amount of S-protein mutations: 15 in the RBD alone and 37 overall[183,184]. Concerningly, the effectiveness of vaccines designed based on the Wuhan-Hu-1 spike decreased starkly against omicron[185]. Despite the great sequence divergence of S-protein, select broadly neutralizing antibodies (bnAbs) continue to neutralize omicron by

targeting ultra-conserved epitopes within the receptor-binding motif (RBM) of RBD[186–190]. The continued evolution of SARS-CoV-2 into more infectious strains that evade immune detection illustrates the molecular arms race between pathogens and the human immune system, and why functional immune repertoire interrogation is essential for our protection. A knowledge of conserved epitopes in the RBM can inform immunization strategies and the design of cross-protective vaccines. Additionally, antibody repertoire mining of omicron-infected individuals will yield new therapeutic mAbs that offer treatment options in the absence of effective vaccines. A breadth of epitope targeting is crucial for the development of therapeutic mAb cocktails that remain effective when challenged with an evolving pathogen.

The combination of YSD and Ig-Seq-driven heavy chain discovery identified a greater diversity and number of clonotypes than each method on its own. Different selection methods (e.g, column immobilized antigen vs biotinylated antigen in solution during flow cytometry) may impact which VHs are isolated. Epitope masking through biotin attachment at surface-exposed lysine residues is a known drawback of chemical biotinylation but can be circumvented through the genetic addition of alternative affinity tags. Both Ig-Seq and YSD discovered NTD and RBD-specific neutralizers; however, only YSD identified S2-directed neutralizers. S2 epitope exposure may be greater during YSD selection, as the assay takes place in solution, thereby increasing the chance of spontaneous S1 shedding compared to ECD immobilization during column chromatography. The ability to select multiple light chain partners for many VHs is an

attractive feature of coupling YSD and Ig-Seq, as it increases the pool of manufacturable candidates. Additionally, the modularity of our synthetic biology-inspired discovery method is easily adapted to affinity maturation and directed evolution of antibodies. The ability to store immune repertoire libraries in yeast permits fast and efficient re-assessment of repertoires. This expedience is highly advantageous for initiatives that aim to develop pan-coronavirus neutralizing antibodies and vaccines, as revived yeast libraries can easily be selected against multiple antigens[191–193]. In summary, our pipeline is capable of discovering therapeutic candidate mAbs in accelerated time frames from primary immune response repertoires and provides the ability to store, combine, and reassess repertoires, as well as study the adaptive immune response.

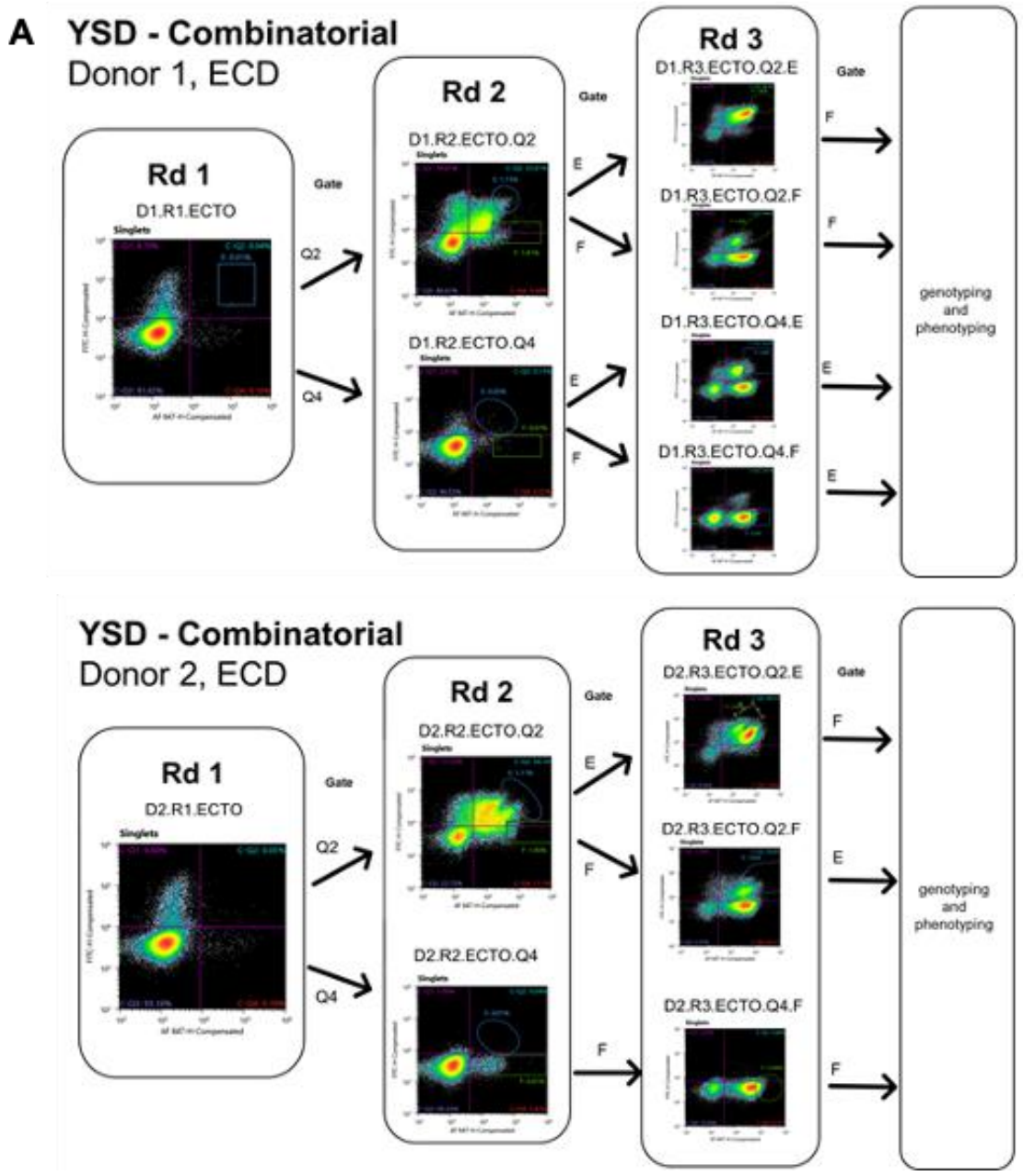

Figure 3.1 YSD selection enriches antigen-specific VH:VLs. **A** Gating strategy over three rounds of antibody selection for Donors 1 and 2. Frequency of specific heavy chain V-gene occurrence in Donor 1's (**B**) and Donor 2's (**C**) repertoire throughout the selection and progressive enrichment of SARS-CoV-2 spike protein-specific heavy V-genes. White indicates no corresponding V-genes were detected. **D** Reduction of heavy V-gene diversity throughout the selection, **E** Tracking of IGHV3-30, -3, and -5 frequency throughout the selection (analysis shown for Donor 2)

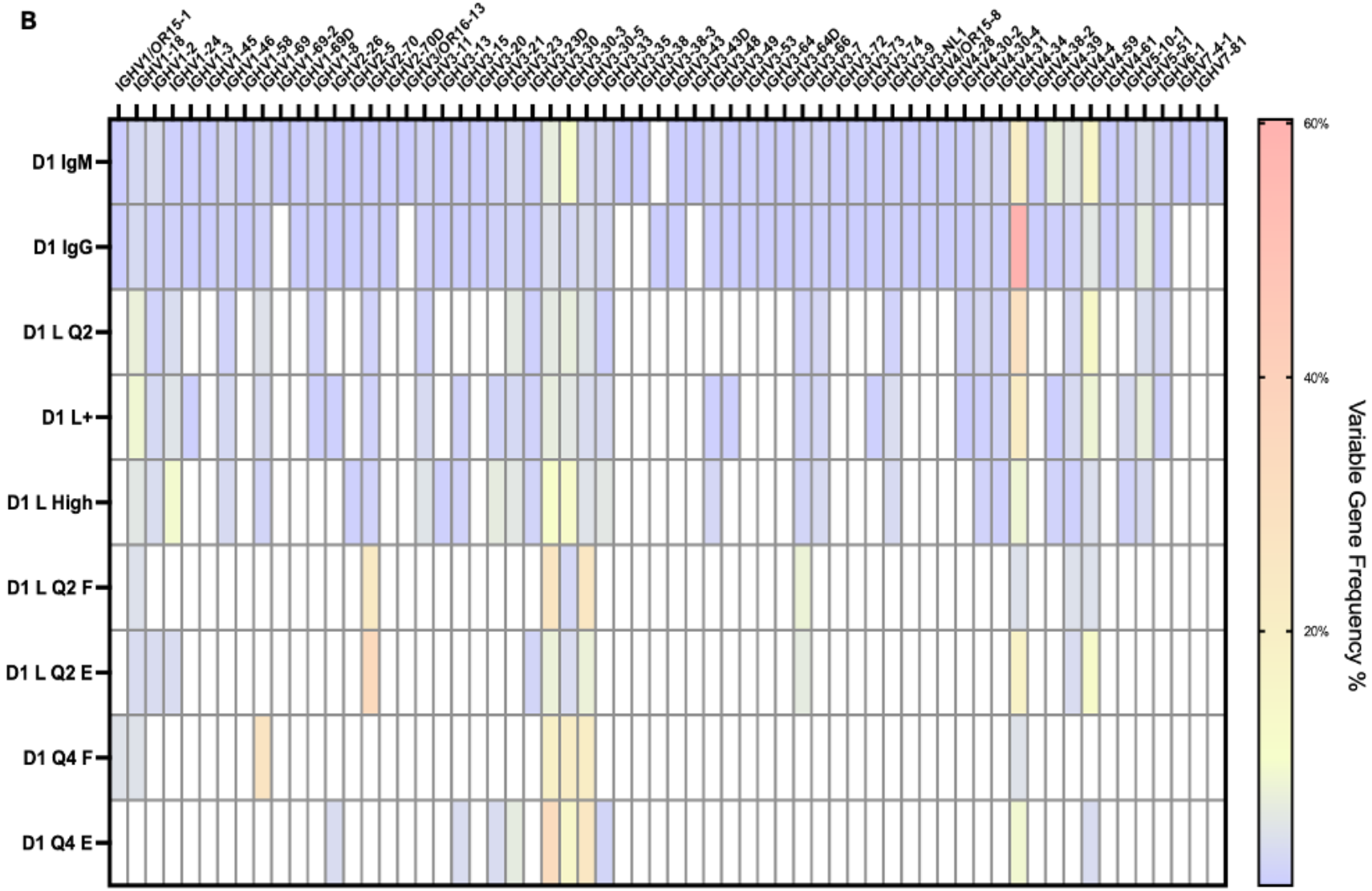

Figure 3.1 (continued). YSD selection enriches antigen-specific VH:VLs.

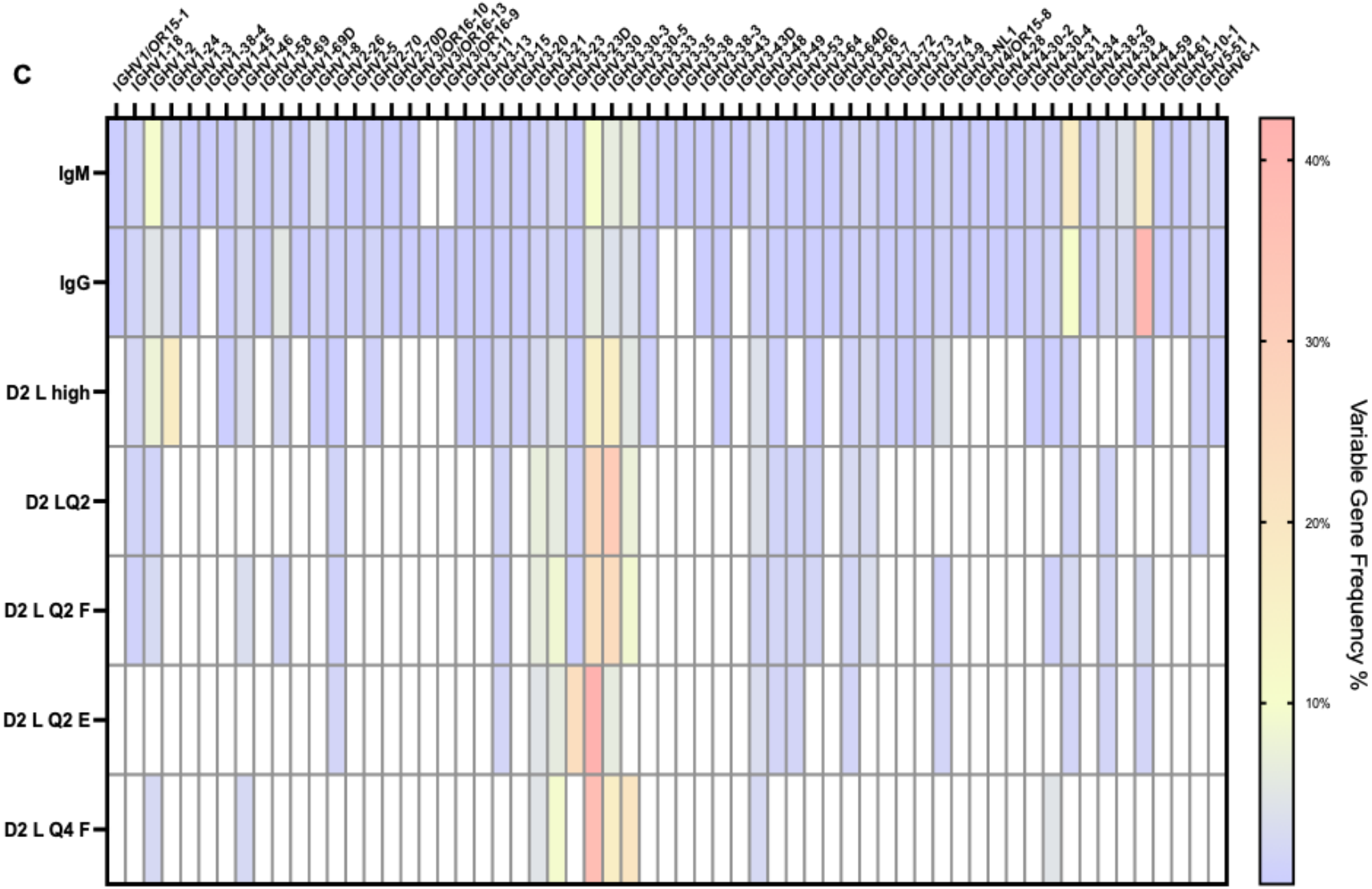

Figure 3.1 (continued). YSD selection enriches antigen-specific VH:VLs.

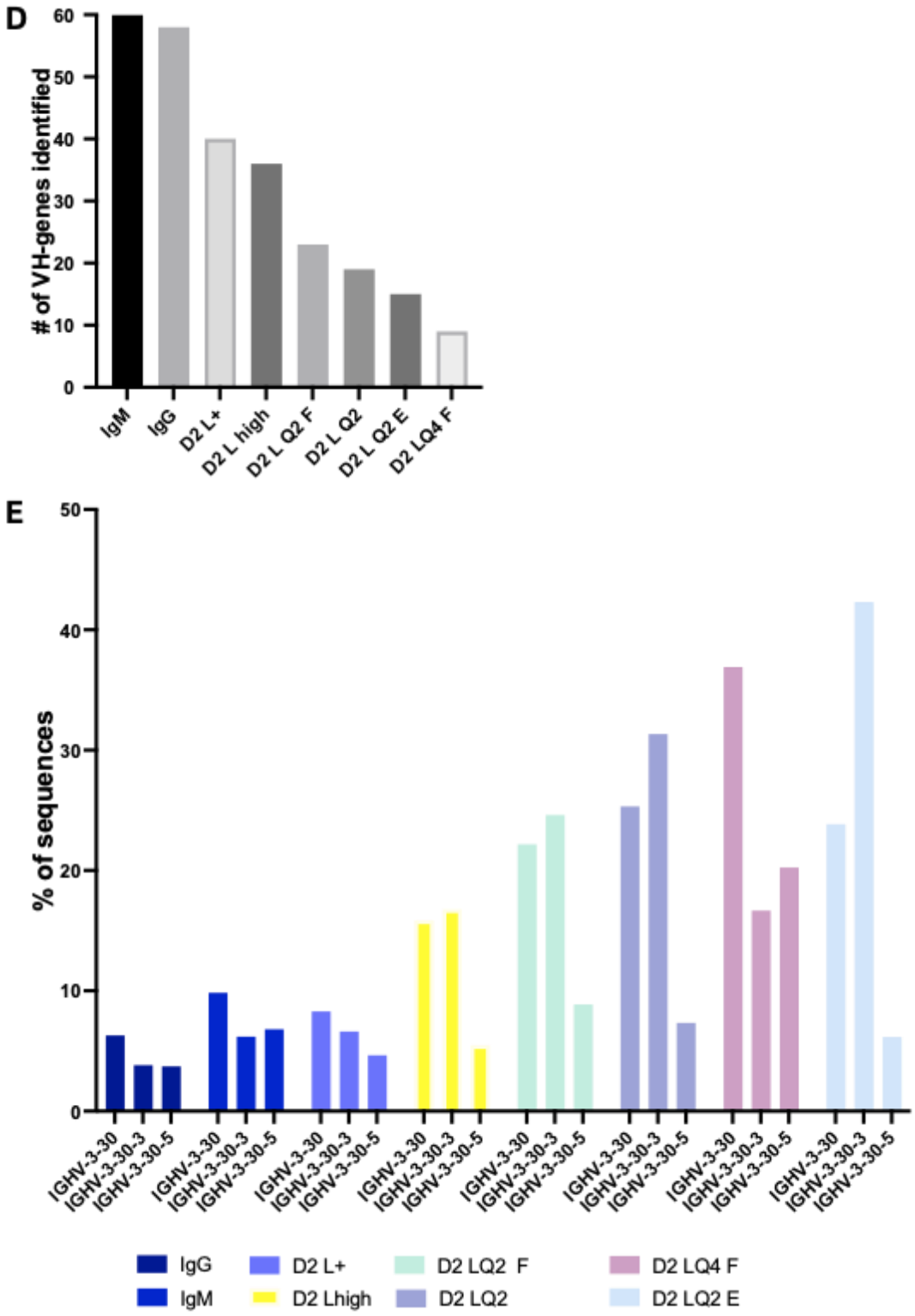

Figure 3.1 (continued). YSD selection enriches antigen-specific VH:VLs.

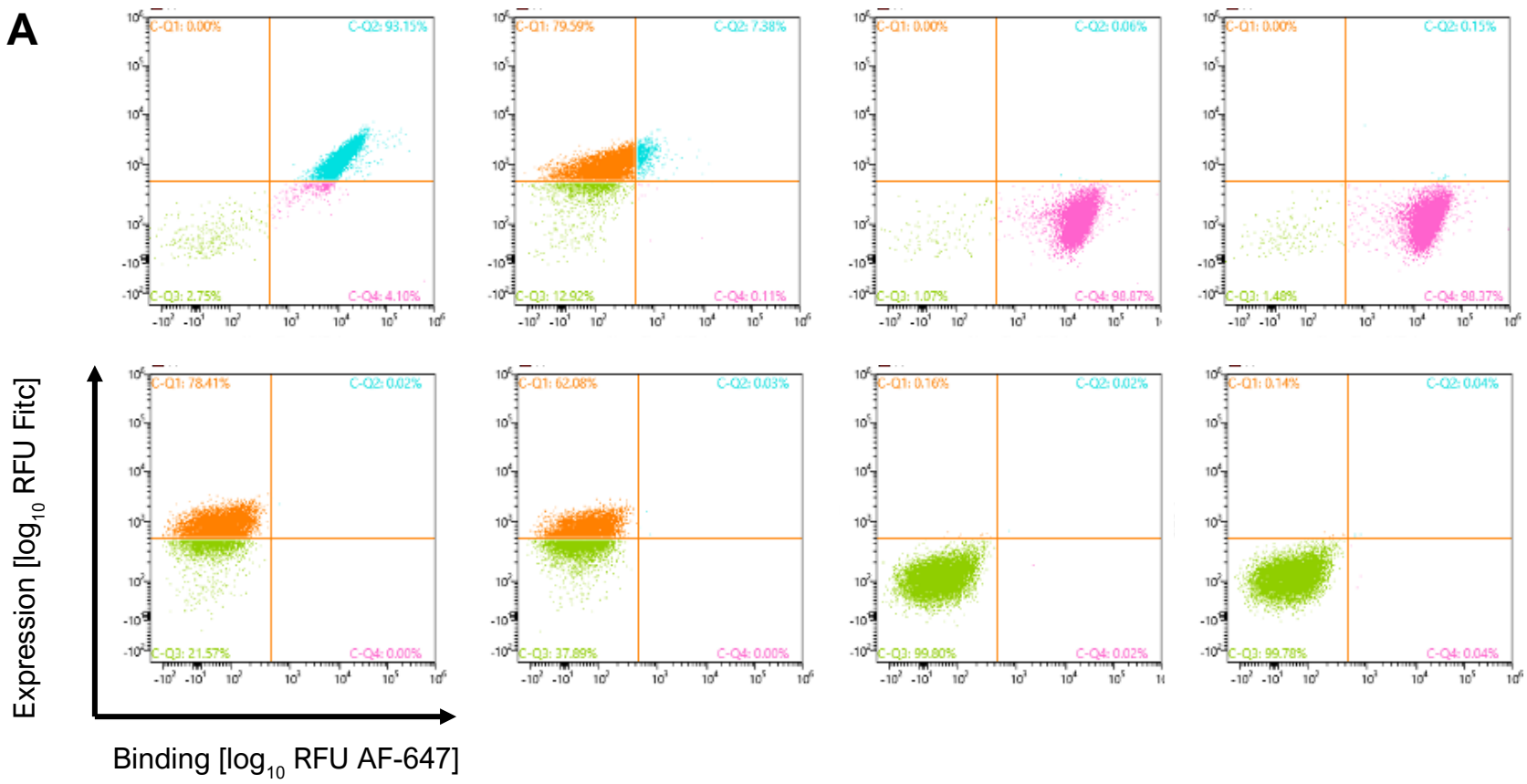

Figure 3.2 Selection of candidate mAbs from YSD. A Spectral analyzer-obtained, representative images of different antibody expression and binding profiles at 200 nM antigen concentration during single-clone phenotyping against the ECD (top row) and RBD (bottom row) of the spike protein. The y-axis shows expression (FITC) while the x-axis shows binding (AF-647). Expression-normalized binding was calculated and plotted as color intensity for every single clone assessed for binding. Clones were picked in duplicate and placed into 96-well plates and then incubated with either the ECD (**B**) or RBD (**C**). A range of Fab-binding intensities and epitope specificities was selected for candidate advancement.

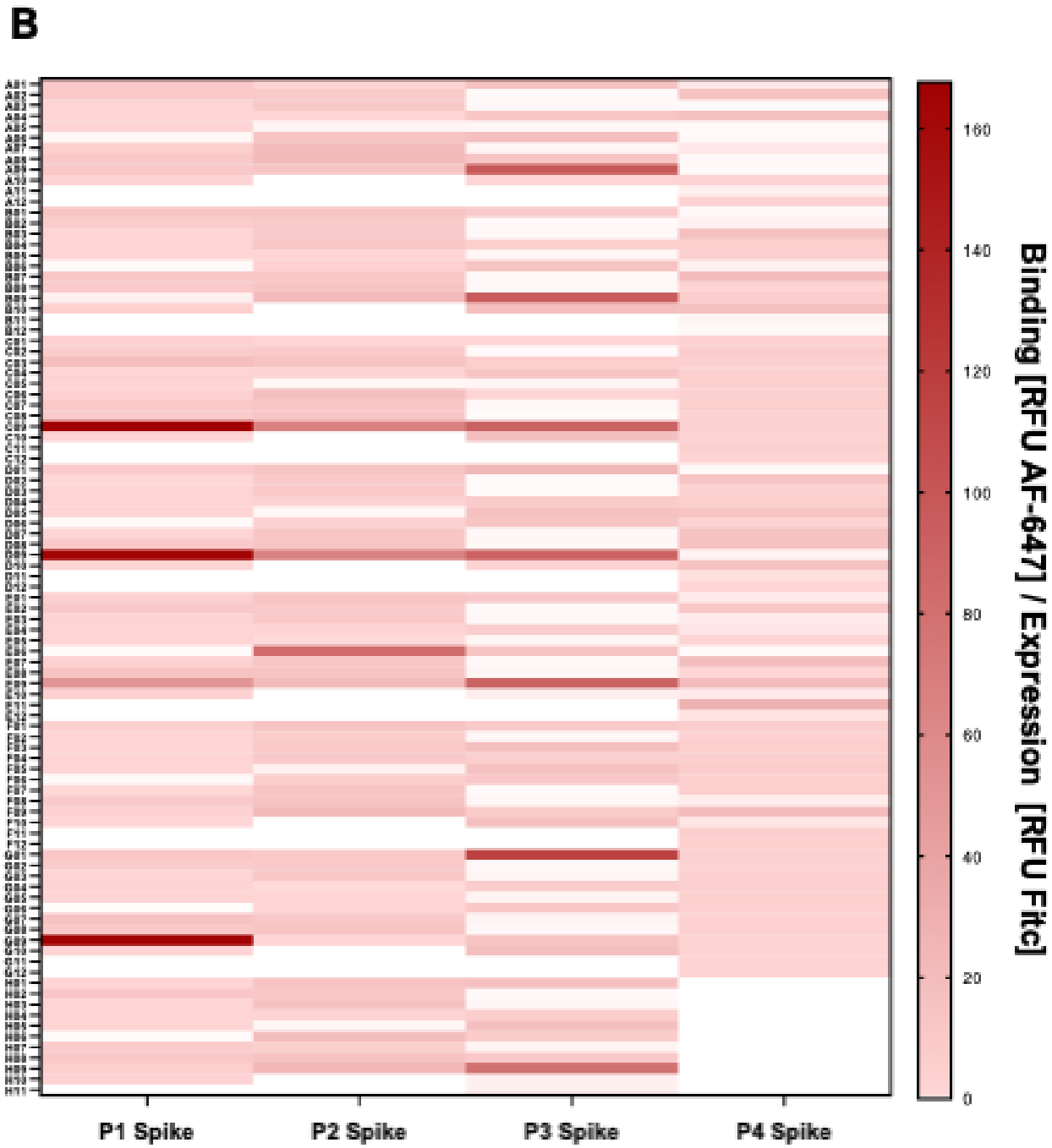

Figure 3.2 (continued). Selection of candidate mAbs from YSD.

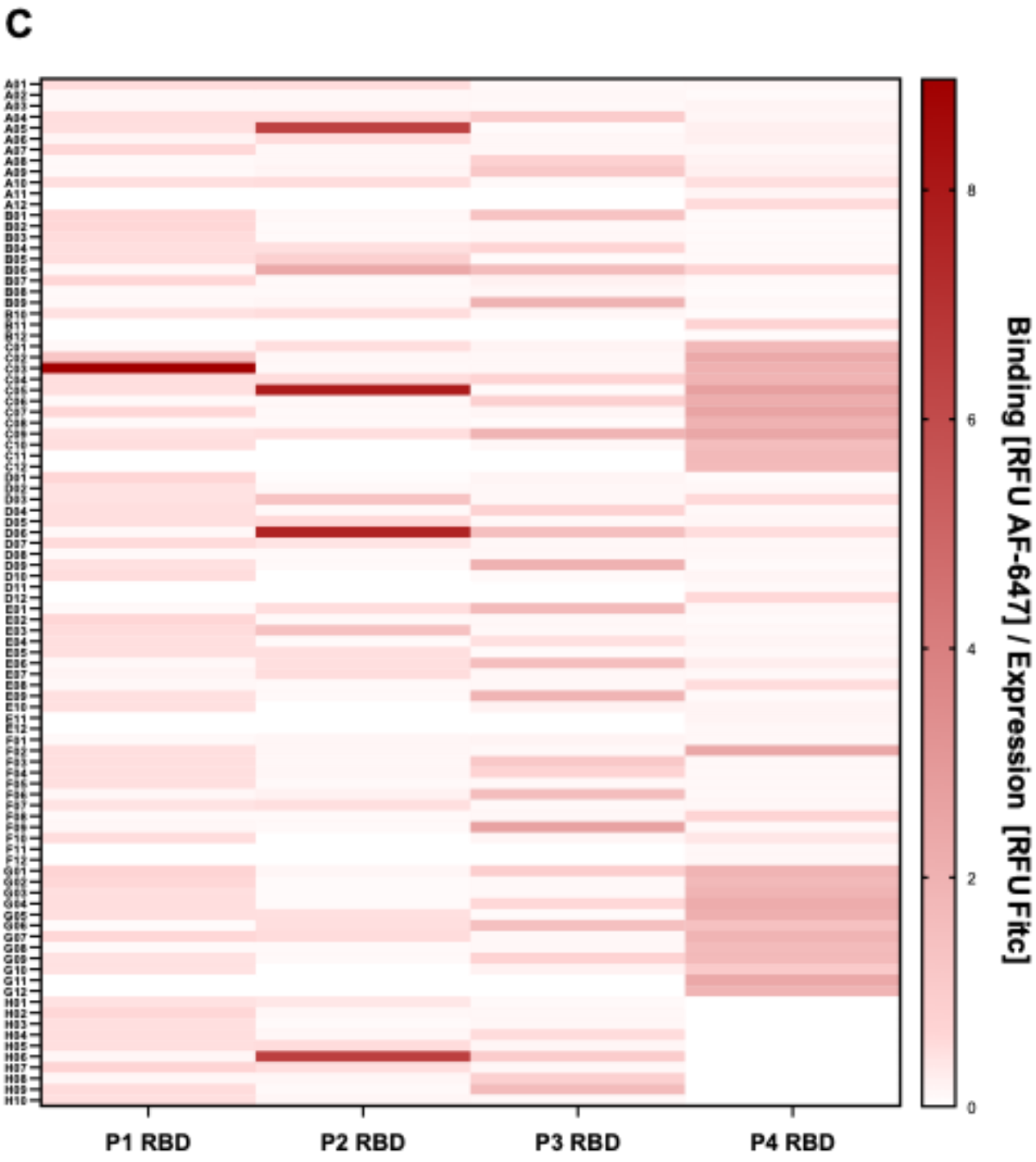

Figure 3.2 (continued). Selection of candidate mAbs from YSD

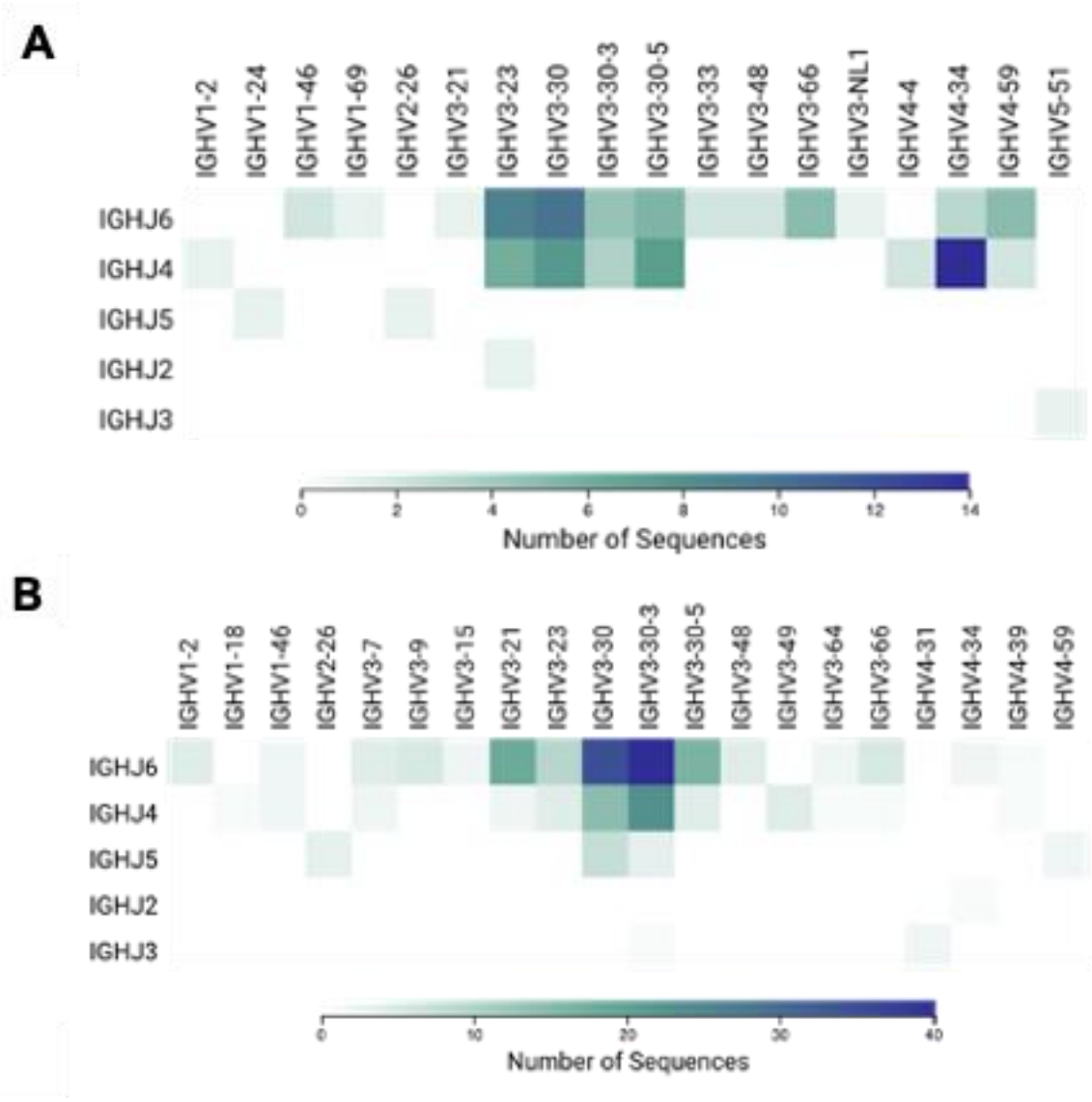

Figure 3.3 Heavy chain V-gene distribution of YSD selected antibodies. A Donor 1 and **B** Donor 2 distribution of IGHVs and J-gene usage as annotated by Geneious Biologics.

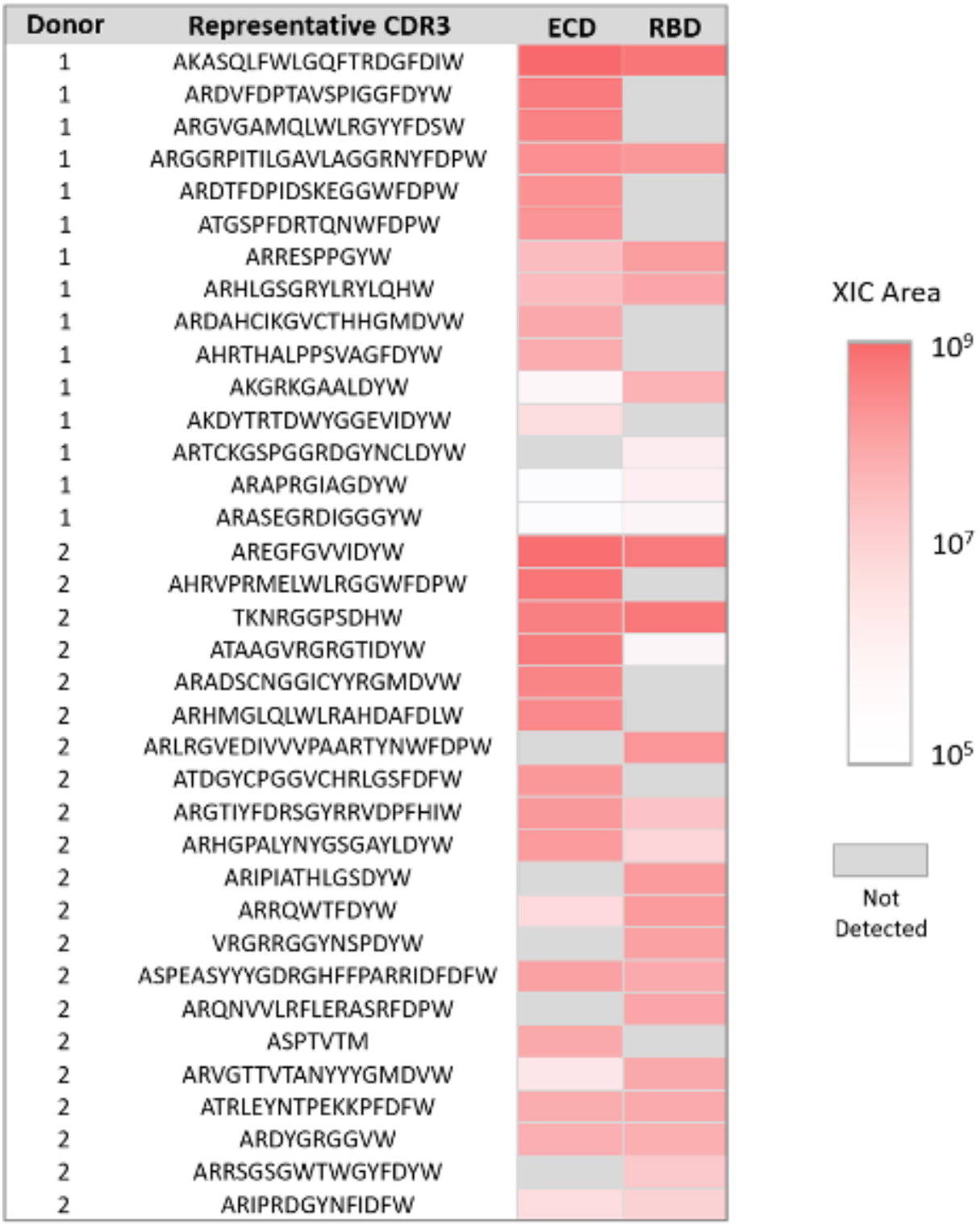

Figure 3.4. Clonotypes identified by Ig-Seq. The area under the peak (XIC) corresponds to the frequency a clonotype is observed at.

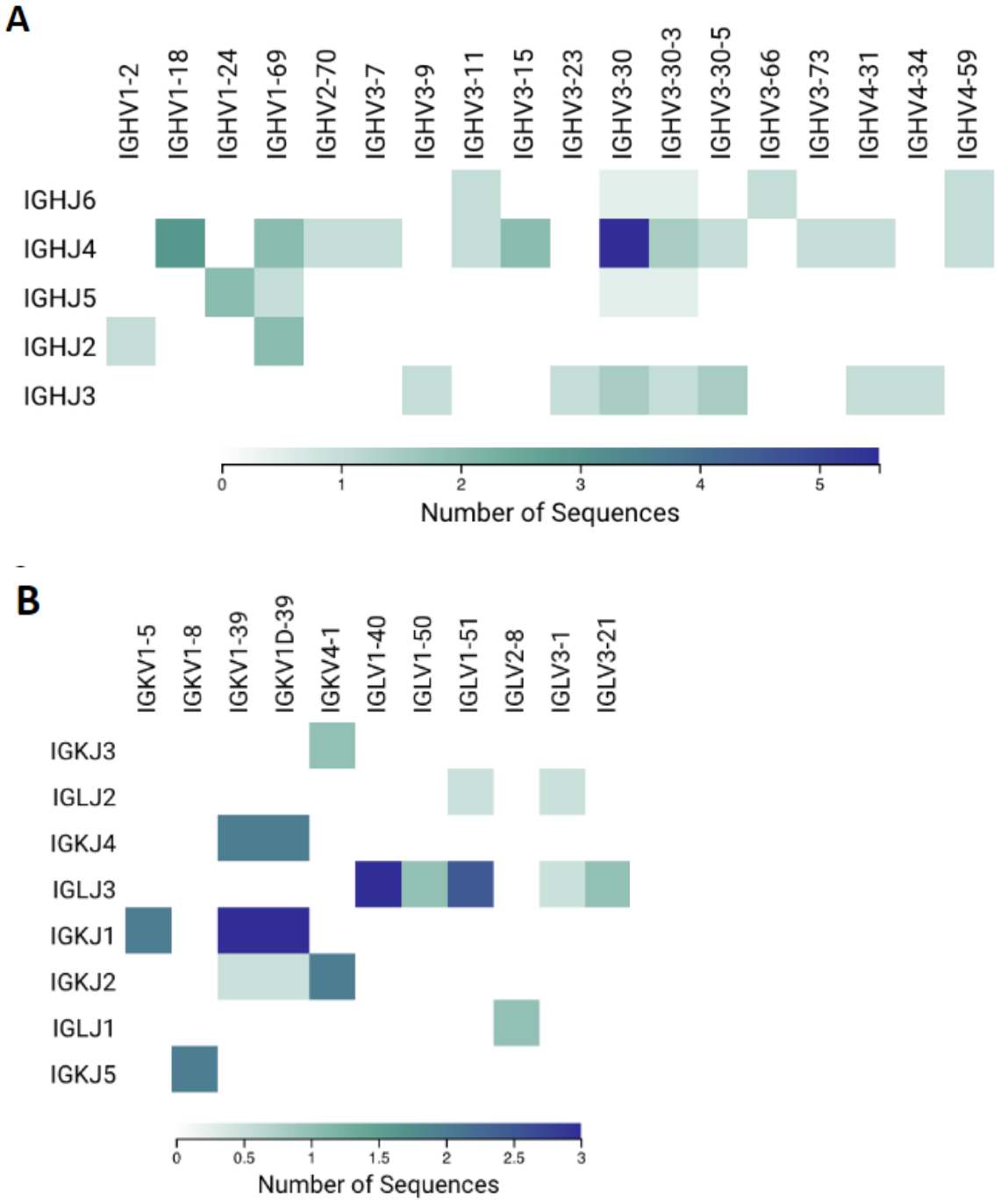

Figure 3.5. VH and VL variable gene distribution of Ig-Seq-YSD samples. **A** Ig-Seq-derived VH sequences display a similar bias towards IGHV3-30 and other IGHV3 family members. **B** YSD identified IGLV1-40, IGKV1-39, and IGLV1-51 as frequent partners for Ig-Seq VHs.

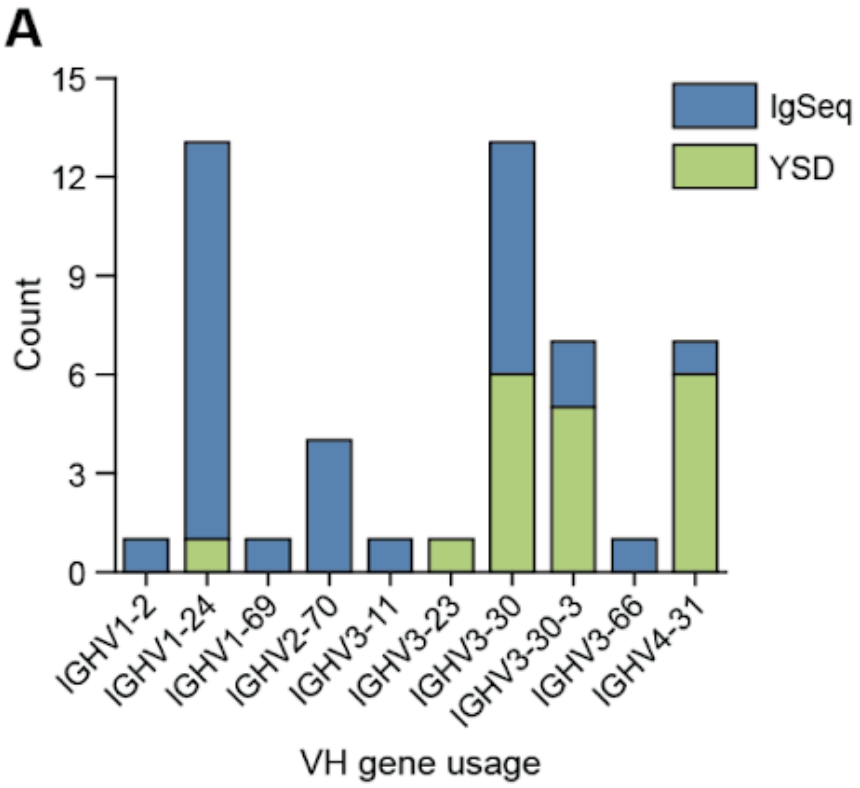

Figure 3.6. IGHV gene usage and IC50 distribution of SARS-CoV-2-neutralizing mAbs. A Neutralizing heavy chain V-gene usage differentiated by discovery method.

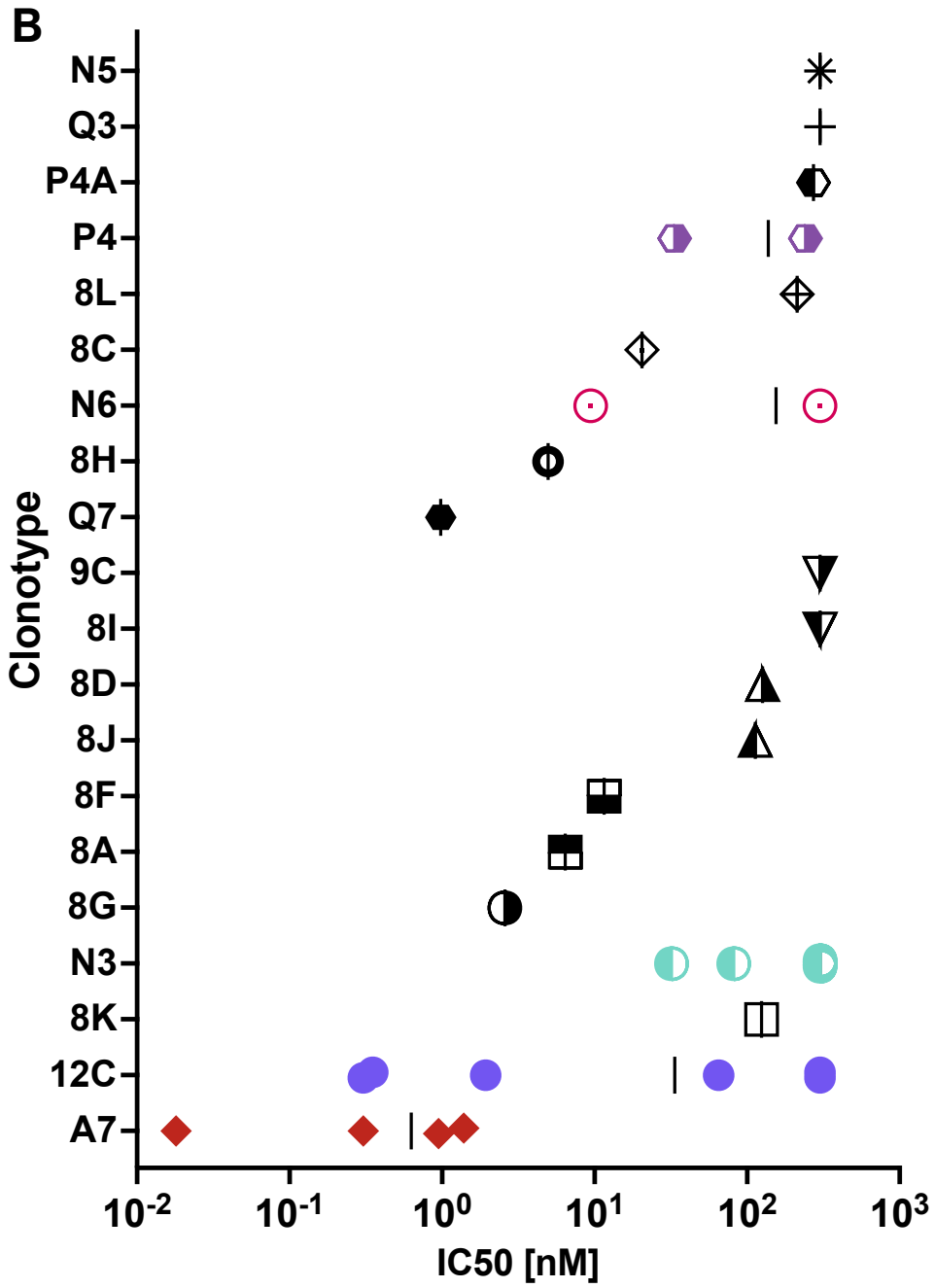

Figure 3.6 (continued). IGHV gen e usage and IC50 distribution of 32 SARS-CoV-2-neutralizing mAbs B IC50 value as determined through curve fitting to authentic live virus neutralization curves. Each unique clonotype is plotted on the y-axis. Clonotypes that formed multiple VH:VL pairs are highlighted in color and each value is plotted to represent the spread of neutralizer potency dependent on light chain usage.

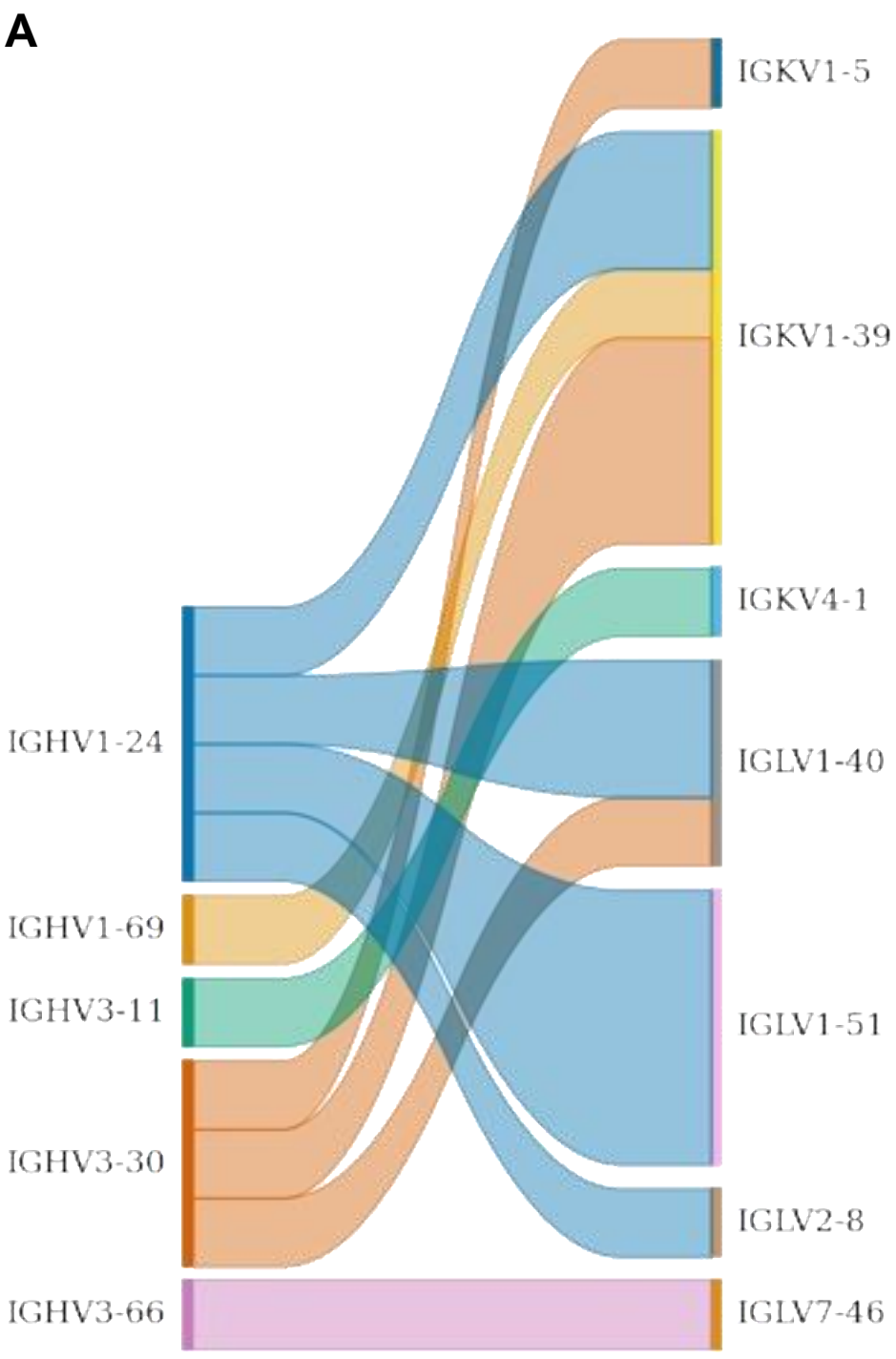

Figure 3.7 VH and VL usage and pairing for Ig-Seq and YSD-derived neutralizing VH:VLs. **A** Ig-Seq-derived neutralizing IGHV light chain pairing distribution. **B** YSD-derived neutralizing IGHV light chain pairing distribution.

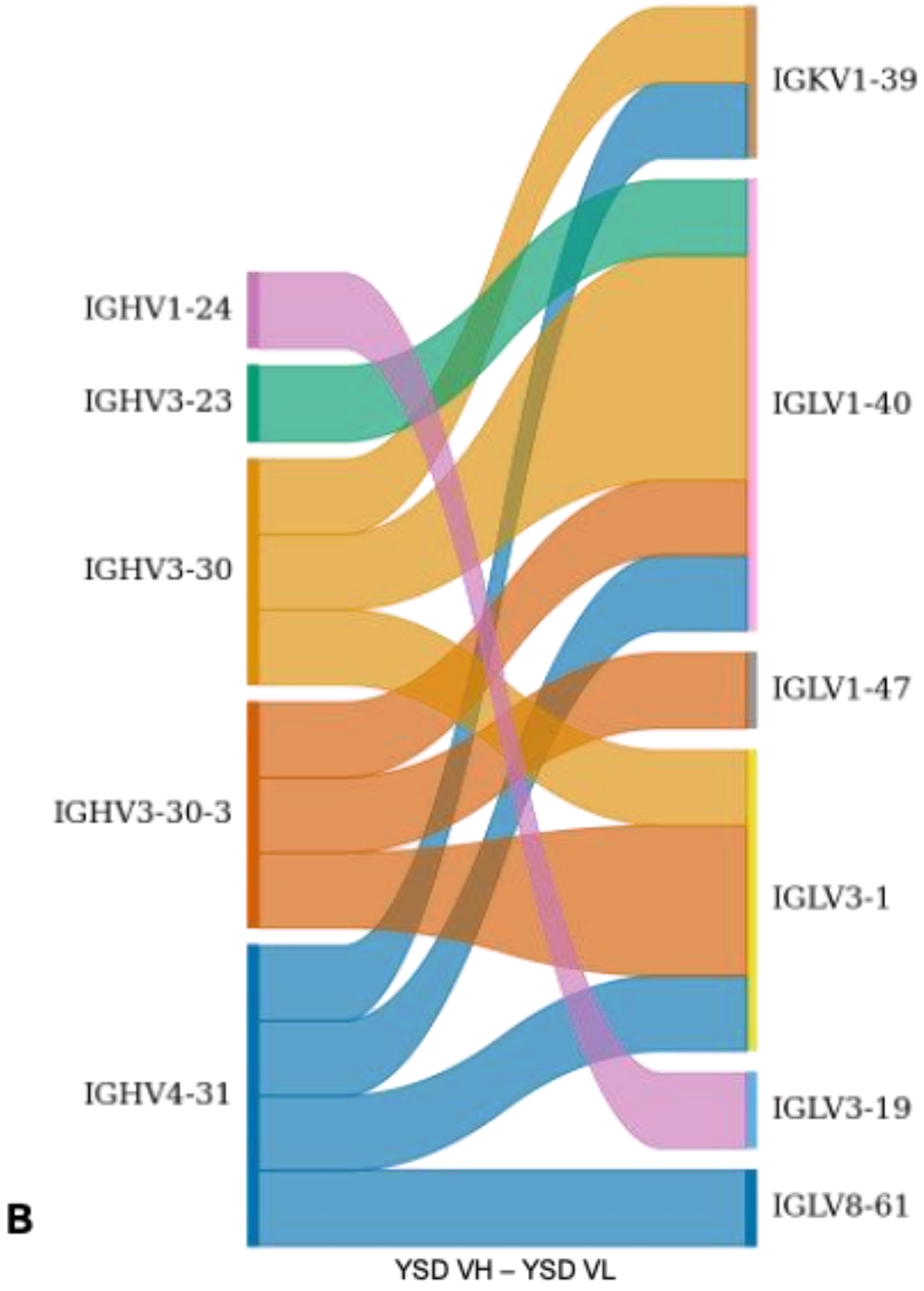

Figure 3.7 (continued). VH and VL usage and pairing for Ig-Seq and YSD-derived neutralizing VH:VLs.

mAb	VH label	IC50 [nM]	VH gene	H-CDR3	VL gene	L-CDR3	Epitope	HC source
A7V32	A7	0.01806	IGHV1-24	ATGSPFDRTQNWFDP	IGLV2-8	SSYAGSNNLA	NTD	IgSeq
A7V131	A7	0.3045	IGHV1-24	ATGSPFDRTQNWFDP	IGLV1-51	GTWDNSLSAGV	NTD	IgSeq
12C114	12C	0.3516	IGHV1-24	ATGPAVRRGSWFDP	IGLV1-51	GTWDSSLSGYV	NTD	IgSeq
A7V3	A7	0.95	IGHV1-24	ATGSPFDRTQNWFDP	IGLV1-51	GTWDSSLSAVV	NTD	IgSeq
Q76	Q7	0.98	IGHV1-24	ATRFAVYGDYLIDY	IGLV3-19	NSRDSSGDLVV	NTD	YSD
A7V132	A7	1.391	IGHV1-24	ATGSPFDRTQNWFDP	IGLV1-51	GTWDSSLSAGV	NTD	IgSeq
12C115	12C	1.929	IGHV1-24	ATGPAVRRGSWFDP	IGLV1-40	QSYDSSLSGWV	NTD	IgSeq
8G91	8G	2.586	IGHV1-69	AREQPPGRMVVPATYWHFDL	IGKV1-39	QQSYSTPYT	RBD	IgSeq
8H97	8H	4.955	IGHV3-30-3	ARGNWGSYYYGMDV	IGLV3-1	QSSDDSNQLV	ECD	YSD
8A3	8A	6.422	IGHV3-66	ARGGVVDTYYYYGMDV	IGLV7-46	LLSQSGAWV	NTD	IgSeq
N6-326	N6	9.42	IGHV3-30-3	ARPYSGSYWGYFDY	IGLV1-47	AAWDDSLSGPV	S2	YSD
8F90	8F	11.56	IGHV3-11	ARDGITNSGVVTHFGMDV	IGKV4-1	QQYNSYSRVT	NTD	IgSeq
8C100	8C	20.45	IGHV3-30	AKDRRYYDFWSGYVGSPYYYYYYGMDV	IGLV3-1	QSYGNNQGV	ECD	YSD
N3-63A	N3	32.0	IGHV4-31	ARGTIYFDRSGYRRVDPFHI	IGLV8-61	TLYMGGGLLV	RBD	YSD
P4D96	P4	33.54	IGHV3-30	AKAPGQWLRFHYYGMDV	IGLV1-40	NSRDINSNHVL	NTD	YSD
12C130	12C	65.13	IGHV1-24	ATGPAVRRGSWFDP	IGKV1-39	QQSYSTPRT	NTD	IgSeq
N3-63B	N3	81.5	IGHV4-31	ARGTIYFDRSGYRRVDPFHI	IGLV1-40	QSYDGSLNDDVI	RBD	YSD
8J127	8J	113	IGHV3-30	AKATQLFWLGQFTRDGFDI	IGLV1-40	QSYDSSLSGWV	RBD	IgSeq
8K42	8K	124.2	IGHV3-30	ARDYGRGGV	IGKV1-39	QQSYSTRPLT	NTD	IgSeq
8D107	8D	126.9	IGHV3-30	AKTSGYNLPDY	IGKV1-39	QQSYSTPRT	ECD	IgSeq
8L63	8L	212.5	IGHV3-30-3	ARAYGGGYLTPFDY	IGLV1-40	NSRDSSGDLVV	ECD	YSD
P4D3	P4	242	IGHV3-30	AKAPGQWLRFHYYGMDV	IGLV1-40	QSYGNNQGV	NTD	YSD
P4A3	P4A	272	IGHV3-30	ARDDTGRVGSGWYCPLY	IGKV1-39	QQSYSTPLS	S2	YSD
8I122	8I	>300	IGHV3-30	AKDRAIFYLNPRYYLDY	IGKV1-5	QQYNSYPLT	ECD	IgSeq
9C-118	9C	>300	IGHV3-30	ASPTVT	IGKV1-39	QQSYSSSWT	ECD	IgSeq
12C112	12C	>300	IGHV1-24	ATGPAVRRGSWFDP	IGLV1-40	QSFDSSLSGPVV	NTD	IgSeq
12C113	12C	>300	IGHV1-24	ATGPAVRRGSWFDP	IGKV1-39	QQSYSTRPLT	NTD	IgSeq
Q318	Q3	>300	IGHV3-30	AKQAGAYCSGGSCYSSSEADY	IGLV1-40	QSYGNNQGV	RBD	YSD
N5C6	N5	>300	IGHV3-23	APGRSLY	IGLV1-40	QSYDSGLSGSI	S2	YSD
N3-819	N3	>300	IGHV4-31	ARGTIYFDRSGYRRVDPFHI	IGKV1-39	QQSYSTRPLT	RBD	YSD
N3-63C	N3	>300	IGHV4-31	ARGTIYFDRSGYRRVDPFHI	IGLV3-1	QAWDSSTVV	RBD	YSD
N6-10B	N6	>300	IGHV3-30-3	ARPYSGSYWGYFDY	IGLV3-1	QAWDSSTF	S2	YSD

Table 3.1 Neutralizing mAbs discovered through Ig-Seq and YSD. The table lists heavy and light V-gene usage of each VH:VL pair, ordered from strongest to weakest IC50 [nM]. Each clonotype has a unique VH label and the last column indicates the discovery arm through which it was isolated.

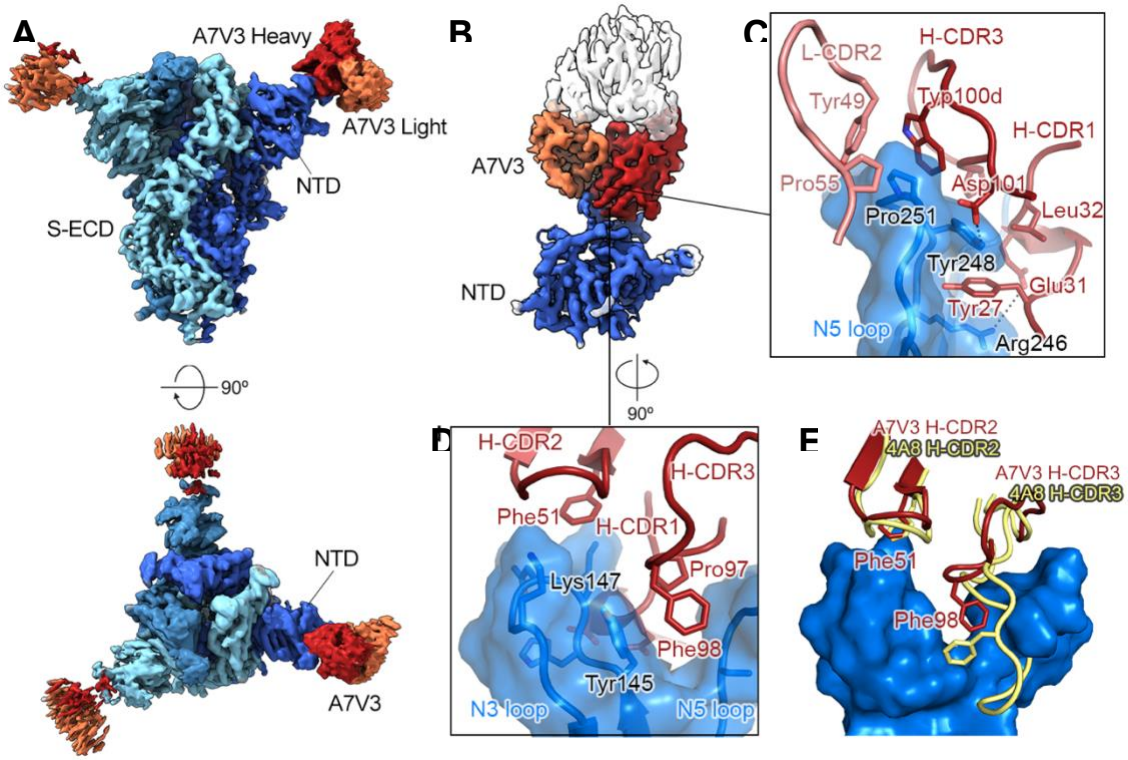

Figure 3.8. A7V3 binds the NTD supersite. **A** Head-on and top-down view of a 3.0 Å map of A7V3 Fab bound to the spike trimer. Protomers are colored steel blue, royal blue, and sky blue. A7V3s VH is colored firebrick and the VH is coral. **B** A7V3 binds the NTD through interactions with the N3 and N5 loops. **C** The CDRH3 and CDRL2 encase the protruding N5 loop through hydrophobic interactions. Heavy chain residues Glu31 and Asp101 form polar interactions with Arg246 and Tyr248. **D** The N3 and N5 loops form a crevice into which CDRH3 Pro97 and Phe98 insert. Phe51 interacts with Lys147 through a pi-cation bond. **E** Overlay of A7V3 bound to the NTD and 4A8 bound NTD (PDB ID: 7C2L) illustrates the conserved Phe51 interaction and the reduced interactions of A7V3's with N3 due to its shorter CDRH3.

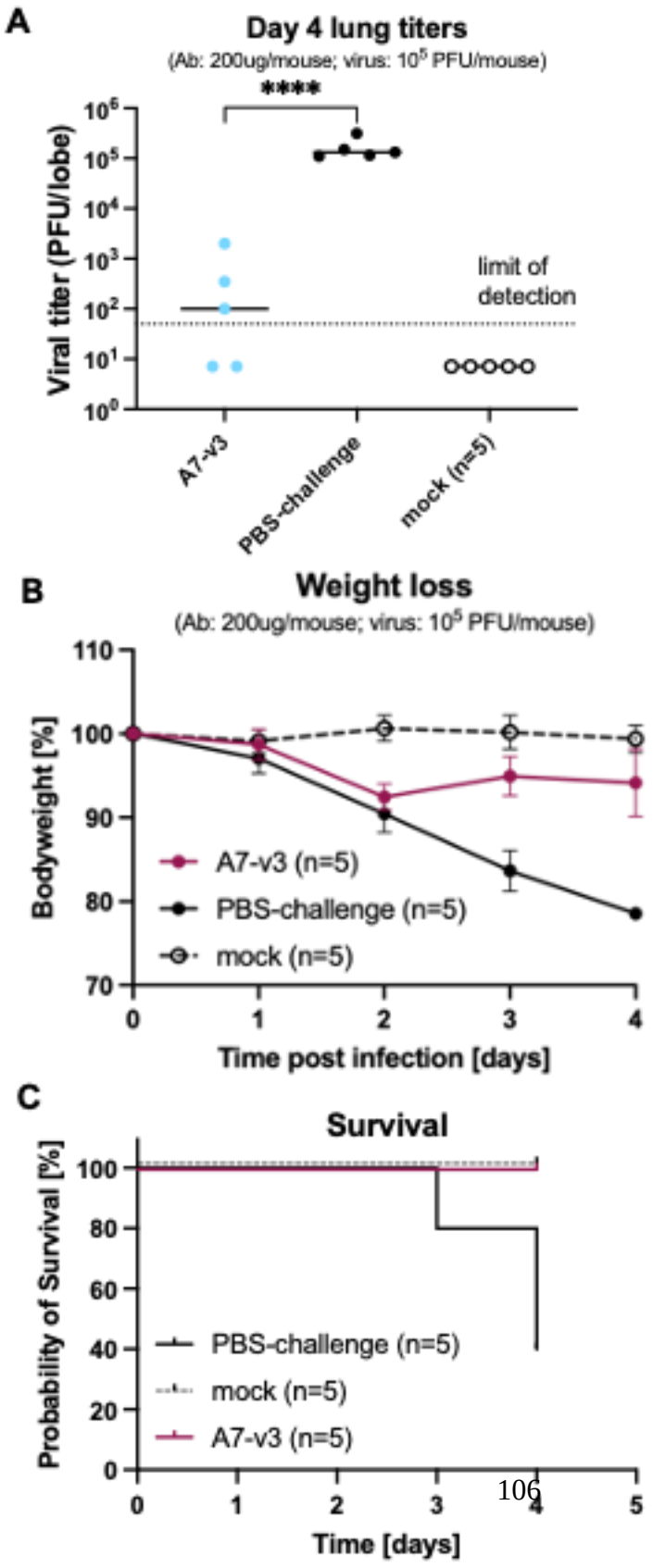

Figure 3.9. Prophylactic administration of A7V3 protects SARS-CoV-2 (MA10) challenged BALB/c mice. **A** Lung titers of mouse-adapted SARS-CoV-2 on day four post challenge with a lethal virus dose (10^5 PFU) in BALB/c mice (n=5) receiving prophylactic A7V3 or PBS control. ****P<0.0001, determined by one-way ANOVA. **B** Weight loss of BALB/c mice (n=5) over four days after viral challenge. **C** Survival of 200µg A7V3 receiving BALB/c mice (n=5) over four days after intranasal MA10 challenge vs PBS control group.

CHAPTER 4: Public Light Chains Expedite Neutralizing Antibody Discovery

4.1 INTRODUCTION

As outlined in section 1.1.2, antibody repertoire diversity initially stems from the combinatorial diversity of V(D)J germline rearrangements, and the pairing of different VH and VL constructs. Following B-cell activation, additional diversity is introduced through SHM. However, antibody repertoire sequencing reveals that only a handful of light chain V-genes encode the majority of VL partners in diverse VH:VL pairings[194]. This bias is not the result of antigen-elicited polarization alone, as evidenced by a comparison of newborn and adult repertoires. Weber et al. observed restricted use and preferential expression of the same Kappa family V-genes in neonates and adults; however, the adult repertoires exhibited antigen-driven sequence diversification due to

---

This chapter has been in part adapted from the manuscript: Jule Goike, Ching-Lin Hsieh, Andrew Horton, Elizabeth C. Gardner, Foteini Bartzoka, Nianshuang Wang, Kamyab Javanmardi, Andrew Herbert, Shawn Abbassi, Rebecca Renberg, Michael J. Johanson, Jose A. Cardona, Thomas Segall-Shapiro, Ling Zhou, Ruth H. Nissly, Abhinay Gontu, Michelle Byrom, Andre C. Maranhao, Anna M. Battenhouse, Varun Gejji, Laura Soto-Sierra, Emma R. Foster, Susan L. Woodard, Zivko L. Nikolov, Jason Lavinder, Will N. Voss, Ankur Annapareddy, Gregory C. Ippolito, Andrew D. Ellington, Edward M. Marcotte, Ilya J. Finkelstein, Randall A. Hughes, James M. Musser, Suresh V. Kuchipudi, Vivek Kapur, George Georgiou, John M. Dye, Daniel R. Boutz, Jason S. McLellan, Jimmy D. Gollihar. Synthetic repertoires derived from convalescent COVID-19 patients enable discovery of SARS-CoV-2 neutralizing antibodies and a novel quaternary binding modality. Biorxiv. doi: https://doi.org/10.1101/2021.04.07.438849

Conceptualization: JG, CH, AH, ECG, DRB, JSM, and JDG; Methodology: JG, CH, AH, ECG, FB, NW, KJ, AH, RR, MJJ, SLW, ZLN, JL, TSS, SVK, VK, RAH, IF, DRB, JSM, and JDG; Investigation: JG, CH, AH, ECG, FB, NW, AA, KJ, AH, WNV, JAC, ACM, RR, MJJ, SLW, ZLN, JL, RAH, SVK, VK, IJF, JMM, DRB, JSM, and JDG; Data Analysis and Interpretation: JG, CH, AH, DRB, NW, AH, JMD, IJF, EMM, JSM, and JDG; Data Curation: JG, CH, AH, ECG, NW, DRB, JSM, AH, JDM, JSM, and JDG; Original Draft: JG, CH, AH, ECG, DRB, JMM, JSM, VK, EMM, and JDG; Review & Editing: JG, CH, AH, ECG, DRB, ADE, EMM, JMM, IJF, GG, JSM, and JDG; Funding: JMD, IJF, GCI, VK, SVK, GG, ADE, JSM and JDG

SMH while the newborn repertoire remained largely germline[195], indicating that light chain repertoire restriction occurs prior to antigen exposure. Additional evidence of a restricted light chain V-gene repertoire is found in CD5+ B-cells, a CD23+ IgD+ IgM+ B-cell subpopulation that produces "natural" (polyreactive, low-affinity IgM) antibodies.[196] The light chain repertoires of CD5+ B-cells are predominantly of germline configuration and demonstrate selective usage of Kappa V-genes, consistent with previously identified preferred V-genes[195]. Interestingly, Foster et al. found that light chain V-gene bias is also present in non-productively rearranged constructs, indicating an underlying molecular mechanism of VJ rearrangement as a possible cause[197].

Light chain repertoires may be more restricted than suggested by theoretical calculations. Observations of a restricted light chain repertoire challenge our existing understanding of the importance of antibody diversity in recognizing a maximal number of pathogens. Though maximal diversity is desirable for the heavy chain repertoire, emerging evidence supports a role for light chain repertoire restriction as an important tool to control antibody autoreactivity[198]. Indeed, an analysis of donor repertoires in myasthenia gravis patients showed that the frequency of light chain V-gene usage varied relative to healthy controls; Vander Heiden et al. reported an increase in IGLV2-14 pairing, and a reduction in the commonly overrepresented V-gene families IGKV1, IGKV4, IGKV2, and IGLV1[199]. Experiments in mouse models of diabetes type I and systemic lupus erythematosus revealed that their antibodies undergo fewer cycles of light chain receptor editing, a main mechanism underlying B-Cell development and self-tolerant antibody repertoire formation[200].

The natural restriction of the light chain repertoire can be harnessed in antibody discovery campaigns. In 2018, Harris et al. created a transgenic rat (OmniFlic rat) with a normal human IGH locus and a light chain locus that only expressed IGKV3-15 with JK1. Their goal was to build a discovery platform for bispecific antibodies with two VH chains of differing antigen specificity but identical light chains. They demonstrated that the OmniFlic rat successfully produced antigen-specific antibodies upon immunization with different proteins[201].

Initial studies of light chain repertoire diversity often relied on single-cell sequencing methods and therefore low n-numbers were assessed. With recent advances in system-wide repertoire sequencing[202], light chain V-gene distribution and frequency data can be mined more meaningfully. In this chapter, I evaluate light chain V-gene distribution in a large VH:VL data set, construct public light chains (PLCs), and use PLCs for antibody pair discovery against SARS-CoV-2.

4.2 RESULTS

4.2.1 Human Light Chain Variable Gene Usage Is Highly Polarized

In 2015, DeKosky et al. published a VH:VL dataset of memory B-cells collected from three healthy donors[203]. Using a high-throughput flow-focusing device, they emulsified single B-cells and performed overlap extension RT-PCR to link the native VH and VL transcripts[202]. I performed a secondary analysis on the 100450 available light chain junctions from this dataset to reveal the frequency at which light chain variable genes successfully pair with a heavy chain. As the donors were not experiencing an

antigenic challenge at the point of blood draw, I did not expect this repertoire to show polarization of VH and VL genes towards a specific antigen. When evaluating the Kappa and Lambda variable gene distribution within the dataset, I found that 54835 VH:VL pairs (54.59%) used a Kappa V-gene and 45615 (45.41%) used Lambda (**Figure 4.1 A**). The human Kappa locus is located on chromosome 2 and is split into a proximal and distal V-gene cluster with a roughly 800 kb-gap separating the clusters (**Figure 4.1 B**)[204]. The V-gene cluster closest to the J and C genes is the proximal cluster. The distal cluster is located further upstream towards the 5' end of the chromosome. The proximal and distal V-genes are duplicates of each other and the distal cluster is oriented as the mirror image of the proximal, apart from the four most 3' V-genes, which are not duplicated and are therefore only present in the proximal locus. Though the proximal and distal sequences of the same V-gene are highly homologous, most can be differentiated through their single nucleotide polymorphisms. Of the 54835 VH:VL Kappa pairs, only 2.84% used a distal Kappa V-gene (**Figure 4.1 C**). Of 44 functional Kappa V-genes and open reading frames (ORFs), 6 genes and 2 alleles were identical in sequence and could therefore not be reliably annotated as either distal or proximal (see **Table 4.1**). However, when identical proximal and distal genes were excluded from analysis, the proportion of distal V-genes increased to 3.7%. The repertoire showed a strong bias towards the use of the proximal Kappa V-genes (72.9%), which appeared independent of protein sequence, considering the protein display between proximal and distal V-genes was largely identical. The remaining 23.24% of Kappa V-genes used in VH:VL pairs was derived from the four non-duplicated V-genes of the proximal cluster (**Figure 1C**). These four V-

genes occurred eight times more frequently than all distal V-genes combined. **Figure 4.2** shows how many times each V-gene was detected in the dataset as a function of its genomic location, illustrating the predominance of proximal V-genes.

Examination of the frequency at which each Kappa V-gene occurs in a VH:VL pair revealed preferential use of eight IGKVs. 18.5% of Kappa VH:VL pairs used IGKV3-20 across the three donors, followed by IGKV1-39 (12.6%), and IGKV1-5 (11.5%) (**Figure 4.3 A**). This distribution was comparable to those reported in previous studies examining significantly smaller datasets acquired through single-cell sequencing[197,205]**.** Interestingly, similar distributions were observed when ethnically diverse donor samples comparing the IGKV repertoire of Australian and Papua New Guinean donors were analyzed[206]. Additionally, IGKV4-1, IGKV3-15, IGKV3-11, IGKV2-28, and IGKV1-33 were frequent partners in the Kappa VL junctions. IGKV3-20, 3-15, and 3-11 were the most frequent proximal IGKV genes with non-identical distal copies. Comparison showed that distal copies were observed one to two orders of magnitude less often. For example, IGKV3-20 was identified 10061 times but IGKV3D-20 only 274 times (**Figure 4.3 B**). This raises interesting questions about the role of distance in V(D)J rearrangement.

Three Lambda V-genes, IGLV1-44, IGLV1-51, and IGLV3-1 were frequently observed in the Lambda V-gene distribution. As the listed VL-genes can pair promiscuously with VH genes, I next constructed germline sequences of IGKV1-5, IGKV2-28, IGKV3-11, IGKV3-15, IGKV3-20, IGLV1-44, IGLV1-51, and IGLV3-1. The most common J-gene usage of a V-gene was derived from the publicly available

iRepertoire data of two donors' light chain repertoires[207–210]. The resulting nine PLCs are listed in **Table 4.2**. The Kappa PLC V-genes constituted 61% of the 54835 Kappa light chain junctions and the three Lambda PLCs were present in 22% of the Lambda junctions (**Figures 4.3 A, C**). Together, the nine Kappa and Lambda PLC V-genes accounted for 43.03% of the 100450 sequences (**Figure 4.3 D**).

4.2.2 PLCs Form Functional Pairs with Anti-SARS-CoV-2 Heavy Chains

The prevalence of the PLC V-genes in VH:VL pairs led us to hypothesize that PLCs could be used to identify productive light chain partners for antigen-specific heavy chains. We tested this hypothesis using SARS-CoV-2 specific VHs largely identified from Ig-Seq. Donor serum antibodies were purified through Protein A column chromatography and affinity purified using the SARS-CoV-2 Wuhan-Hu-1 spike protein. The recovered antibodies were then digested with trypsin for mass spectrometry analysis. During digestion, VH:VL pairs become uncoupled and light chain information is lost.

To test whether SARS-CoV-2-specific VHs could generate functional antibodies with PLCs, we paired each VH with each of the PLCs and produced the full-length IgG. The nine IgGs of each VH were screened for spike protein binding through ELISAs. In total, we paired 22 VHs with 9 PLCs to test 198 IgGs and determined their EC50s (**Figure 4.4 A**). To test for polyreactive binding, the VH:PLC pairs were screened against the spike ectodomain and its' RBD and NTD subunits, the RBD, and NTD. A given VH should only show reactivity towards two of the three at most, given that the NTD and RBD subunits have no overlapping epitopes.

Encouragingly, almost every VH was capable of forming a functional binder with one or multiple PLCs. Fifteen of the VH:PLC pairs had EC50s under 10 nM, with VH N6 and PLC 2 (N6-2) and VH N3 and PLC1 (N3-1) reaching sub-nanomolar values of 560 and 190 pM, respectively. Twenty VH:PLC pairs had EC50s between 10 and 100 nM (**Figure 4.4 A**). No VH generated a productive pair with every PLC. The greatest promiscuity was observed with B3, which successfully paired with five PLCs, producing high-affinity mAbs in the narrow range of 1.08 to 4.66 nM. Other VHs showed a broad spread of affinities when paired with different PLCs. Most notably, N3 showed the tightest binding at 0.19 nM with all VH:PLC pairs as N3-1, 1.65 and 1.23 nM with PLC3 and 7, 99.63 and 76.34 nM with PLC5 and PLC9, and very weak affinity with PLC2. In contrast, N6 displayed an exclusive preference for PLC2. Only three VHs (9C, 2C, and 8B) did not form~~ed no~~ robust PLC pairs, although residual binding was detected with some PLC combinations (**Figure 4.4 A**). Whether this is due to a lower binding affinity of the VHs towards the spike protein or unsuitability between the VH and the PLCs is uncertain. However, 9C with PLC 5 (9C-5) exhibited low levels of binding activity against the ectodomain at the highest concentrations measured, but 9C-5 did not show any binding against the RBD subunit, indicating that ECD binding is likely not caused by polyreactivity (**Figure 4.4 B**).

The most potent binders identified through ELISA screening were submitted for neutralization testing of the live SARS-CoV-2 virus. Seventeen VH:PLC pairs neutralized SARS-CoV-2. Ten of these targeted the RBD, five the NTD, and two the S2 subunit (**Figure 4.5 A, Table 4.3**). The IC50 values ranged from 0.25 nM for N3-1 to

values exceeding 300 nM for the RBD binders 4A7 and 1D5. PLC 1, 5, 7, and 8 each occurred three times in VH:PLC pair neutralization and only PLC6 failed to produce a neutralizing mAb (**Figure 4.5 B**). N3 neutralized SARS-CoV-2 in combination with PLCs 1, 3, and 7.

N3 light chain partners were also selected through YSD. Other than most PLC paired VHs, N3 was originally derived from YSD selections. However, the initial round of selection failed to produce an optimal light chain partner for N3, perhaps because N3 contains an endogenous AaRI cut site that reduces the probability of its correct assembly into antibody expression vectors. After identifying N3 as a highly potent binder through PLC screening, a codon-optimized version was paired with donor light chain repertoires and the YSD selection was repeated. Interestingly, the YSD selection did not yield an improved light chain partner for N3. In fact, PLC1 and 3 were the most potently neutralizing N3 pairings with PLC7 in fourth place (**Figure 4.6**). YSD did identify four additional neutralizing light chain partners for N3; however, only one was superior to N3 with PLC7. Only the third most potent mAb used a rare light chain V-gene with IGLV8-61; the remaining three non-PLC neutralizers used frequently observed V-genes, including IGKV1-39 and IGLV1-40. N3-63C used IGLV3-1, the same IGLV as PLC9, with minor mutations to the germline sequence, suggesting that weak VH:PLC pairings could be promising starting points for affinity maturation efforts.

4.2.3 N3-1 Binds a Unique Quaternary RBD Epitope and PLC1 Participates in Binding

To understand how VH:PLC pairs recognize their epitopes and neutralize SARS-CoV-2, we performed CryoEM on N3-1, B3.1, and N6-2 in complex with the stabilized Wuhan-Hu-1 spike protein. Based on the previous ELISA epitope mapping, N3-1 was expected to bind the RBD and B3.1 and N6-2 the S2 subunit.

The negative EM stain of N3-1 in complex with the spike trimer suggested that two Fab fragments of a single IgG molecule bind the three RBD subunits simultaneously **(Figure 4.7**). This binding mode is further substantiated by surface plasmon resonance assays comparing the affinity of N3-1 Fabs and full-length IgG (**Figures 4.8 A, B**): the apparent affinity of N3-1 IgG is almost 1000-fold enhanced over the Fab and their dissociation constants differ remarkably, with IgG showing minimal dissociation over time relative to Fab.

The 2.8 A global resolution cryo-EM structure of N3-1 Fabs bound to the spike trimer revealed one Fab engaging one RBD subunit in its up conformation and a second Fab binding the remaining two RBD subunits: one in the up and the other in the down conformation at the same time (**Figure 4.9 A**). Focused refinement of the dually engaged Fab enhanced the interpretability of the density in this region and allowed us to distinguish two entirely separate epitopes on the RBD-up and RBD-down (**Figure 4.9 C**). The RBD-up binding site covers 862 $Å^2$ surface area through contacts made by the Fab CDRH1, CDRH3, and CDRL2. PLC1 is also engaged in binding at the RBD-down interface, where CDRH2, CDRH3, and CDRL3 make contacts over a 696 A2 surface area. The CDRH3 engages both RBD-up and down through hydrophobic and polar

interactions of its 18 aa-long CDRH3 loop. On RBD up, CDRH3 residues Tyr98, Phe99, and Arg100a pack against a hydrophobic pocket that is highly conserved from SARS-CoV to SARS-CoV-2 and is formed by Tyr369, Phe377, Lys378, and Pro384.

This pocket is also targeted by SARS-CoV and SARS-CoV-2 cross-reactive antibodies CR3022 and COVA-1 and is largely inaccessible in the RBD-down conformation[211]. CDRH3 residue Tyr98 interacts with Lys378 on the upper ridge of the pocket through cation- π interactions and likely forms a salt bridge with Asp100H within the CDRH3. CDRH3 residues Phe99 and Arg100a Further increase the strength of the Fab – RBD-up binding site through hydrogen bonds with Cys379 and Tyr369. PLC1's CDRL2 contributes through polar interaction of its Tyr49 and Glu55 residues with a sidechain guanidinium of Arg408, forming a secondary binding site on the RBD-up interface together with CDRH1.

The RBD-down binding site includes eleven residues that are also necessary for ACE2 binding out of nineteen total contacts made (**Figure 4.9 D**). Here, a hydrophobic pocket is formed by N3-1 through CDRL3's Trp96, CDRH2's Tyr52, Tyr58, and CDRH3's Arg100e, and Val100g. Phe486 inserts into this pocket and forms cation-π interactions with Arg100e. Arg100e also shares a hydrogen bond with Tyr489 through its guanidinium side chain. PLC1 is further involved in binding through polar interactions of Ser03 CDRL3 with Asn487 and Gln493, which also involve Thr57 of CDRH2. N3-1 effectively blocks ACE2 binding through steric hinderance at the RBD-up site and pinning the second RBD in the down position, occluding the ACE2 binding site (**Figure 4.9 E**).

In summary, N3-1 forms extensive interactions with neighboring RBDs in a quaternary epitope, establishing a novel binding modality in which both the CDRL2 and CDRL3 of PLC1 actively contribute to binding.

4.2.4 N3-1 Binds Robustly to Spike Proteins from Emerging Variants of Concern

As N3 was isolated from donor repertoires challenged with the original Wuhan-Hu-1 strain, we needed to ascertain whether N3-1 could neutralize emerging variants of concern (VOC) with spike protein mutations Genomic surveillance of SARS-CoV-2 variants by Houston Methodist identified twelve emerging mutations in early 2021[160]. Loss of N3-1 binding activity was not observed for any of the twelve mutations by ELISA. More concerning variants diminished the potency of the therapeutic antibodies REGN10987 and REGN10987. To assess the accumulated mutations in isolation, we used mammalian surface display of the single amino acid mutation-containing proteins. In contrast to the Regeneron antibodies, N3-1 was not negatively affected by ~~all~~ the twelve mutations and showed marginally improved binding to L455A, G476D, E406W, Q493F, K444A, K444Q, V445A, G446A, G446V, and the Australian VOC S477N under spike display assay conditions. Mutation F486K caused a minor 6% reduction in binding and Y453A measurements did not differ from WT control (**Figure 4.10 B**). SPR measurements of N3-1 binding to spike variants B.1.1.7 and B.1.351 established an apparent affinity of 296 pM for B.1.1.7 (Alpha) and 300 pm for B.1.351 (Beta), which are slight reductions in affinity compared to 67.9 pM for Wuhan-Hu-1 (**Figure 4.11 A**).

In two independent live virus neutralization tests, N3-1 neutralized all common VOCs: alpha, beta, gamma, and delta (**Figures 4.11 C, D**).

4.2.5 N6-2 and B3.1 Bind Related Epitopes of the S2 Subunit

N6-2 and B3.1 are two neutralizing S2 binders that share a largely identical heavy chain sequence but possess different CDR3s (see **Table 4.3**). To elucidate their binding mode, we solved cryo-EM structures of N6-2 and B3.1 Fabs bound to the S2 devoid of its stalk domain (ΔstalkS2). Both target epitopes at the S2 apex, which may only become exposed and available for antibody binding once S1 shedding has occurred (**Figure 4.12 A**). Interestingly, the CDRH2 loops of N6-2 and B3.1 share the sequence ISYDGSNK and tuck into the same binding pocket at the stalk apex but partly form different contacts (**Figures 4.12 B, C**). The interaction is mediated through hydrogen bonds of CDRH2 Tyr52(A) with Thr761 and the peptide backbone of Gly757. In N6-2, Tyr52A's ring is positioned optimally for π interactions with the downward angled Leu753, forming the hydrophobic ceiling of the binding pocket. In B3.1, Tyr52 is angled more sharply but still maintains the hydrogen bond with Thr761. B3.1's Ser5 forms a hydrogen bond with Asn764, while in N6-1 Ser55 is twisted to face away from the S2 and Asn56 instead interacts with Asn764. Additionally, N6-2 Asp53 contacts Thr739. The structural positioning of N6-2's CDRH2 in the S2 pocket results in a higher number of its residues being involved in S2 binding compared to B3.1.

B3.1 and N6-2 CDRH3s bind different but overlapping epitopes of the S2 domain. B3.1 Arg96 forms a salt bridge with Glu748 and hydrogen bonds between

Tyr100B and Asp745 and Gly97 and Ser750. In contrast, N6-2 engages Glu748 through a hydrogen bond with Tyr96. CDRH3 Ser97 and Trp100A form hydrogen bonds with the backbone of Ser746, and Tyr100 contacts Asp737 (**Figures 14.3 A, B**).

In B3.1, the only light chain residue involved in S2 binding is Tyr91 through a hydrogen bond with Gln853. Examination of the structure reveals an important role for PLC1 in stabilizing and positioning CDRH3 to facilitate interactions with its epitope. Light chain residues Asn92, Trp96, Tyr36, and Glu55 form contacts with CDRH3 residues Tyr100A, Tyr100, Met100D, and Arg96, creating a supportive arch for CDRH3. The CDRH3 of N6-2 is similarly held in place by a network of contacts with PLC2. Light chain residues Asp34, Tyr36, Ser46, and Tyr96 interact with CDRH3 residues Tyr100C, Phe100D, Gly104, and Tyr100 (**Figures 4.13 C, D**).

The most recently emerged VOC, omicron, has accumulated ten amino acid mutations throughout its S2 domain. Assessment of each substitution through mammalian display identified N856K as the most detrimental mutation for both N6-2 and B3.1 binding (**Figure 4.14 A**). Interestingly, neither antibody forms direct contacts with this residue. Asp856 is located in a loop immediately adjacent to the fusion peptide proximal region (FPPR). The FPPR is a highly flexible region that was first described by Cai et al. and is theorized to enhance stabilization of the RBD-down state[212]. Gln 853 is located in the FPPR site and forms a hydrogen bond with Tyr91 of B3.1's light chain (**Figure 4.14 B**). How N856K prevents antibody binding is unclear: Asn856 has no contacts or interactions with the antibodies or S2 residues involved in antibody binding. However, the distance between Gln853 and the terminal side-chain amide groups of Asn856 is only

6.709 Å **(Figure 4.14 B)**. The lysine side chain contains two additional CH2 – CH2 bonds that possess an average length of 1.54 A. An additional 3.08 Å of side-chain length may be sufficient to reduce the distance between Lys856 and Gln853 for hydrogen bond formation between the carbonyl group of Gln853 and the ammonium of Lys856. This would cause Gln853 to be unavailable for binding by the B3.1 light chain residue Tyr91, and result in the restructuring of the FPPR adjoining flexible loop harboring Asn856 Larger structural rearrangement could affect the neighboring CDRH3 and CDRH2 epitopes, resulting in further loss of binding interactions**.**

N853K is the only mutation of ten in the S2 domain that negatively affects B3.1 binding. Seven mutations improve binding and two show no effect (**Figure 4.14 A**). N764K in isolation causes the strongest improvement in binding for B3.1 and N6-2. N764K is located in the shared CDRH2 binding pocket and interacts with B3.1 Ser55 and N6-2 Asn56. The positively charged ammonium group of the substituted lysine likely causes a stronger interaction with the carbonyl groups of Ser55 and Asn56. Though the majority of mutations improve binding, the net effect of the accumulated mutations is abrogated binding, suggesting that N853K causes a large structural rearrangement that occludes relevant epitopes, such as the CDRH2 binding pocket.

4.3 Materials And Methods

Public light chain screening. We analyzed a previously published dataset[203] of $_{VH}$-VL pairs and chose the top six most abundant kappa and top three most abundant lambda V-genes (**Figure 4.3 A, C**). We then constructed germline versions of these V-

genes with their most frequently observed J gene, as published on the iRepertoire website[207–210] (**Table 4.2**) and obtained full-length light chains for each PLC. SARS-CoV-2 specific Ig-Seq heavy chains were then expressed with each PLC as a full-length IgG1. These public light chains were paired with select heavy chains from Ig-Seq and/or YSD for ELISA screening using recombinant S2P.

Monoclonal antibody expression and purification. VH:VL candidates were cloned into custom Golden Gate cloning-compatible pCDNA3.4 vectors for IgG1 expression. For transfections, VL was mixed 3:1 with a corresponding VH. Plasmids were transfected into Expi293F cells (Invitrogen) using the recommended protocol. Monoclonal antibodies were harvested at 5-7 days post-transfection. Expi293F cells were centrifuged at 300 x g for 5 min, supernatants were collected and centrifuged at 3000 x g for 20 min at 4°C and diluted with 1X PBS to a final concentration. Each supernatant aliquot was passed through a Protein G or A agarose affinity column (Thermo Scientific). Flow-through was collected and passed through the column three times. Columns were washed with 10 CV of PBS and antibodies were eluted with 5-mL 100 mM glycine, pH 2.5 directly in neutralization buffer containing 500-µL 1 M Tris-HCl, pH 8.0.

ELISA. Antigen ELISA plates were made using high-binding plates (Corning, catalog no. 3366) with antigen diluted in PBS to a final concentration of 2 µg/mL. Antigen solution (50 µL) was added to microplates and incubated overnight at 4°C with shaking at 100 rpm, 3 mm orbit. Plates were treated with PBSM (2% milk in PBS) blocking buffer at RT for 1 h. Plates were washed three times with 300 µL PBS-T (0.1% Tween-20). Purified antibodies were prepared to 10 µg/mL in PBSM and serially diluted.

Plates were incubated with antibody for 1 h at RT. Plates were washed three times with PBS-T before incubation with goat anti-human Fab-HRP (Sigma-Aldrich, cat. No. A0293) secondary antibody 50 μL) at 1:5000 in PBSM at RT for 45 min. HRP substrate (50 μL) was added to wells and the reaction proceeded for 5-15 min until quenched with 50-μL 4 M H_2SO_4. Absorbance was analyzed at 450 nm ~~in~~ using a plate reader.

Live virus neutralization assays. A SARS-CoV-2 microneutralization assay was adapted from an assay used to study Ebola virus[213]. This assay also used SARS-CoV-2 strain WA1. Antibodies were diluted in cell culture medium in triplicate. A SARS-CoV-2 monoclonal antibody was used as a positive control. An antibody that does not bind SARS-CoV-2A was used as a negative control. Diluted antibodies were mixed with the SARS-CoV-2 WA1 strain, incubated at 37°C for 1 h, and added to Vero-E6 cells at a target MOI of 0.4. Unbound virus was removed after a 1 h incubation at 37°C and culture medium was added. Cells were fixed 24 h post-infection and the number of infected cells was determined using a SARS-CoV-S specific mAb (Sino Biological, cat. no. 401430-R001) and a fluorescently labeled secondary antibody. The percent of infected cells was determined with an Operetta high-content imaging system (PerkinElmer) and Harmonia software. Percent neutralization for each monoclonal antibody at each dilution was determined relative to untreated, virus-only control wells.

Live virus neutralization (VN) for variants of concern B.1.1.7 and WA/1 were performed in an orthogonal assay. The ability of the monoclonal antibodies to neutralize SARS-CoV-2 was determined with a traditional VN assay using SARS-CoV-2 strain USA-WA1/2020 (NR-52281-BEI resources), as previously described[214–217]. All

experiments with SARS-CoV-2 were performed in the Eva J Pell ABSL-3 laboratory at Penn State and were approved by the Penn State Institutional Biosafety Committee (IBC # 48625). For each mAb, a series of 12 two-fold serial dilutions were assessed from a stock concentration of 1 mg/mL. Triplicate wells were used for each antibody dilution. One-hundred tissue culture infective dose 50 (TCID50) units of SARS-CoV-2 were added to two-fold dilutions of the diluted mAb. After 1 h incubation at 37°C, the virus and mAb mixture was added to Vero E6 cells (ATCC CRL-1586) in a 96-well microtiter plate and further incubated at 37°C. After 3 days, cells were stained for 1 h with crystal violet-formaldehyde stain (0.013% crystal violet, 2.5% ethanol, and 10% formaldehyde in 0.01 M PBS). The endpoint of the microneutralization assay was determined as the highest mAb dilution at which all three (or two of three) wells were not protected from virus infection. The percent neutralization for each mAb dilution was calculated based on the number of wells protected: 3, 2, 1, 0 of 3 wells protected was expressed as 100%, 66.6%, 33.3%, or 0%.

Surface plasmon resonance. To investigate the kinetics of mAb N3-1 binding to the spike proteins, purified His-tagged spike variants (SARS-CoV Tor2 strain S-2P ectodomain, SARS-CoV-2 Wuhan-Hu-1 strain S-HexaPro variant, SARS-CoV-2 B.1.1.7 S-Hexapro and SARS-CoV-2 B.1.351 S-HexaPro) were immobilized on a Ni-NTA sensor chip (GE Healthcare) using a Biacore X100 (GE Healthcare). For Fab binding experiments, we immobilized spike proteins to a level of ~450 response units (RUs). Serial dilutions of purified Fab N3-1 were injected at concentrations ranging from 400 to 6.25 nM over spike-immobilized flow cell and the control flow cell in a running buffer

composed of 10 mM HEPES pH 8.0, 150 mM NaCl and 0.05% Tween 20 (HBS-T). Between each cycle, the sensor chip was regenerated with 0.35 M EDTA and 50 mM NaOH, followed by 0.5 mM $NiCl_2$. For IgG-binding experiments, spike immobilization of 200 RUs was used to avoid mass transport effects. Serial dilutions of purified IgG N3-1 were injected at concentrations ranging from 25 to 1.56 nM over a spike-immobilized flow cell and the control flow cell. For the SARS-CoV Tor2 S-2P binding experiments, IgG N3-1 concentrations ranging from 100 to 6.25 nM were used. Response curves were double-reference subtracted and fit to a 1:1 binding model or heterogeneous ligand binding model using Biacore X100 Evaluation Software (GE Healthcare).

Negative stain EM for spike-IgG complexes. To investigate mAb N3-1 binding to spike proteins, purified SARS-CoV-2 Wuhan-Hu-1 S-HexaPro was incubated with a 1.2-fold molar excess of IgG N3-1 in 2 mM Tris pH 8.0, 200 mM NaCl, and 0.02% NaN_3 on ice for 10 min. The spike-IgG complexes were at a concentration of 0.05 mg/mL in 2 mM Tris pH 8.0, 200 mM NaCl, and 0.02% NaN_3 prior to deposition on a CF-400-CU grid (Electron Microscopy Sciences) that was plasma cleaned for 30 s in a Solarus 950 plasma cleaner (Gatan) with a 4:1 ratio of O_2/H_2 and stained using methylamine tungstate (Nanoprobes). Grids were imaged at a magnification of 92,000X (corresponding to a calibrated pixel size of 1.63 Å/pix) in a Talos F200C TEM microscope equipped with a Ceta 16M detector. The CTF-estimation, particle picking, and 2D classification were performed in cisTEM[218]**.**

Cryo-EM sample preparation and data collection. Purified SARS-CoV-2 S (HexaPro variant) was incubated at 0.2 mg/mL with 5-fold molar excess of Fab N3-1 in 2

mM Tris pH 8.0, 200 mM NaCl, and 0.02% NaN_3 at RT for 30 min. The sample was then deposited on plasma-cleaned UltrAuFoil 1.2/1.3 grids before being blotted for 4 s with -3 force in a Vitrobot Mark IV and plunge-frozen into liquid ethane. A 0.2 mg/mL concentration of purified SARS-CoV-2 S with three RBDs covalently trapped in the down conformations (HexaPro-RBD-down variant, S383C/D985C[219–221]) was complexed with a 2-fold molar excess of Fab A7V3b and deposited on plasma-cleaned UltrAuFoil 1.2/1.3 grids before being blotted for 3 s with -4 force in a Vitrobot Mark IV and plunge-frozen into liquid ethane. For the HexaPro-N3-1 sample, 3,203 micrographs were collected from a single grid. Data were collected at a magnification of 22,500X, corresponding to a calibrated pixel size of 1.07 Å/pix.

Cryo-EM data processing. Gain reference- and motion-corrected micrographs processed Warp[171] were imported into cryoSPARC v2.15.0[172], which was used to perform CTF correction, micrograph curation, particle picking, and particle curation via iterative rounds of 2D classification. The final global reconstructions were then obtained via *ab initio* reconstruction, iterative rounds of heterogeneous refinement, and subsequent non-uniform homogeneous refinement of final classes with C1 symmetry. Given the low occupancy of the Fabs on a trimeric spike, C1 symmetry was used for the final runs of heterogeneous and homogeneous refinement. To better resolve the Fab-spike interfaces, both datasets were subjected to particle subtraction and focused refinement. Finally, both global and focused maps were sharpened using DeepEMhancer[173]. For N3-1-RBD model building, we used one RBD-up and one RBD-down from S-HexaPro (PDB ID: 6XKL[222]) and two homologous Fab structures (PDB ID: 5BV7 and 5ITB) as an initial model to

build into map density via UCSF ChimeraX. Both models were further built and iteratively refined using a combination of Coot[176], Phenix[177], and ISOLDE[178].

4.4 DISCUSSION

Given the expansive combinatorial diversity of the human antibody repertoire, identifying suitable VH:VL pairs has been a bottleneck for high-throughput techniques. Here, I show that a small subset of promiscuous VLs, identified from natively paired donor repertoires, is capable of pairing with non-cognate VHs and generating highly potent and specific mAbs. We observed a range of functional contributions from the light chains in VH:PLC paired antibodies. In N6-2, the light chain executes a purely structural role that enables optimal positioning of the VH to bind the S2. The VL makes no contacts with the antigen and is oriented away from it. Conversely, PLC1 takes an active role in epitope recognition when paired with N3. N3-1 is the most potent RBD binder discovered from the SARS-CoV-2 donor repertoires and binds ~~the~~ RBD distinctively. B3 pairs with other PLCs and these pairs have very similar EC50s, but N3-1 is substantially more potent than other N3 and PLC or YSD pairs. This disparity might be explained by the requirement for PLC1 to participate extensively in antigen binding in N3-1. B3.1 only benefits from binding contribution from PLC1 residue Tyr91, but the complex quaternary epitope of N3-1 requires a network of binding interactions from PLC1, which may completely or only partially be supplied by the alternative PLC and YSD light chain partners. As the native light chain partner of N3 is unknown, its cognate partner may also use IGKV1-5; however, it is not guaranteed that native pairs exigently are combinations

of superior quality. V-gene rearrangement during B-cell development's purpose is to produce a stable, non-self-reactive BCR; this process is entirely antigen-agnostic and therefore does not generate an optimal binder for a specific target. Thus, VH:VL pairing is subject to a degree of randomness, and a different VH:VL combination may bind a theoretical antigen with higher affinity. Though N3 likely has undergone affinity maturation, which mutates both the heavy and light chain to create higher affinity binders, its best partner was PLC1, a germline configuration of IGKV1-5, and YSD screening failed to identify a better partner from SMH-experienced donor light chain repertoires. The promiscuity of B3 suggests that multiple VLs can be "good enough" to form stable, antigen-specific pairs. This knowledge can be helpful during therapeutic lead development or to provide alternatives if manufacturability assessments are unfavorable for a specific pair. Future experiments validating the set of nine PLCs against non-SARS-CoV-2 antigens are needed to establish whether PLCs can be used as a tool for antibody pairing independent of antigen specificity of the VH. Whether initial VH:PLC pairings can be improved through targeted mutagenesis to expand their utility is a promising area of future research.

In summary, PLCs enable diverse modes of antigen binding and potentially powerful tools to expedite antibody discovery.

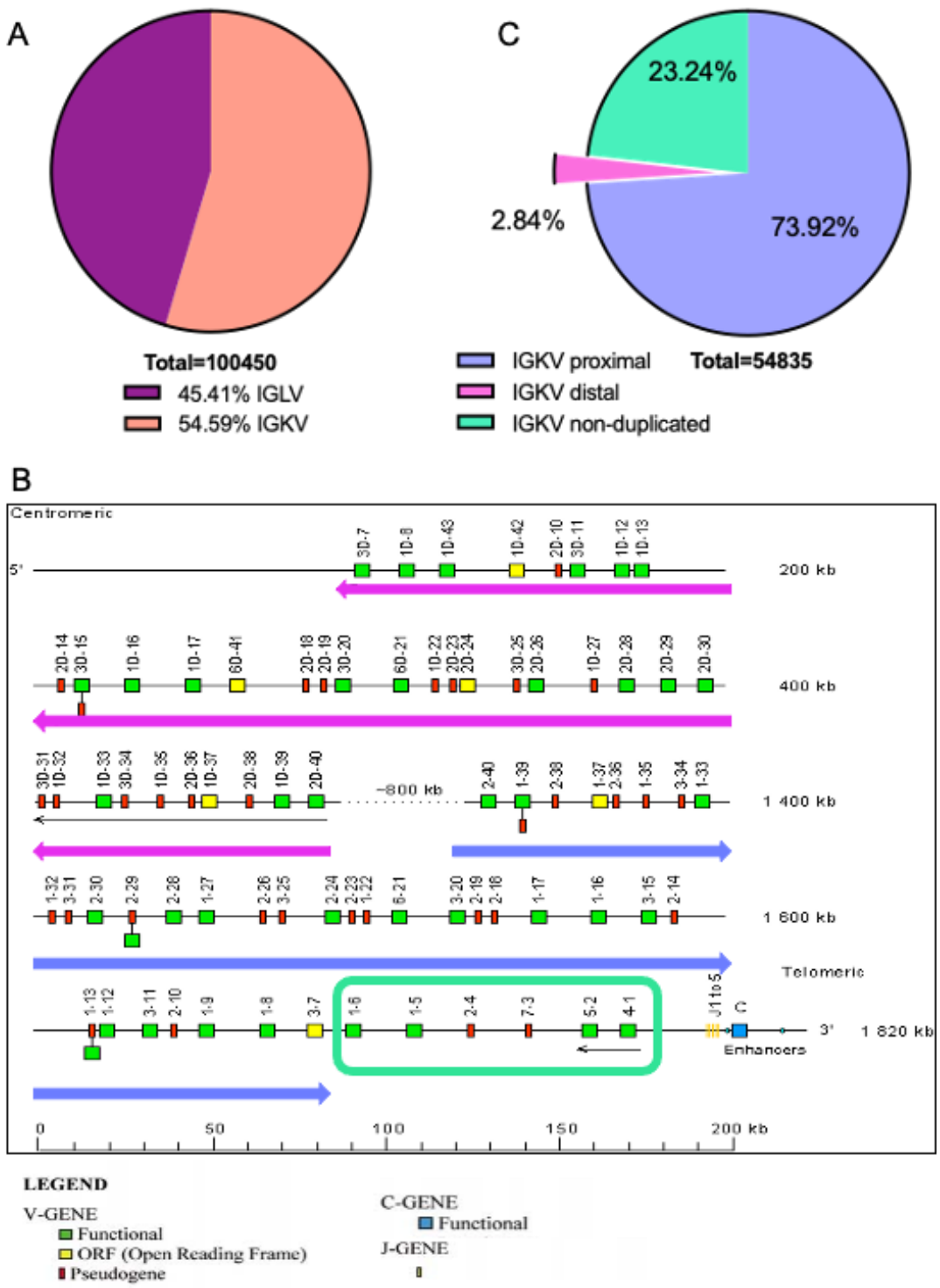

Figure 4.1. Donor light chain repertoires use Kappa proximal V-genes preferentially. A shows the distribution of Kappa and Lambda V-genes 100450 light chain junctions of three donors. **B-C**, Magenta arrows indicate the distal locus and its

coding direction. Blue arrows indicate the proximal locus, and the cyan box denotes the non-duplicated V-genes. The map was adapted from LeFranc, 2001[204]. Though the proximal and distal Kappa loci largely possess identical protein displays, the proximal locus accounts for the majority of paired Kappa light chains. The four non-duplicated proximal V-genes are observed 8 times more frequently than the 22 distal V-genes combined.

Proximal locus V-gene	Distal locus V-gene
IGKV6-21*01	IGKV6D-21*01
IGKV1-33*01	IGKV1D-33*01
IGKV1-39*01	IGKV1D-39*01
IGKV1-12*02	**IGKV1D-12*02**
IGKV1-37*01	IGKV1D-37*01
IGKV2-40*01	IGKV2D-40*01
IGKV2-28*01	IGKV2D-28*01
IGKV1-13*02	**IGKV1D-13*02**

Table 4.1. Kappa proximal and distal V-genes with an identical nucleotide sequence. These V-genes are indistinguishable during annotation because of their identical sequence. Bolded font signifies proximal and distal pairs that differ by a single nucleotide. * denotes the allele.

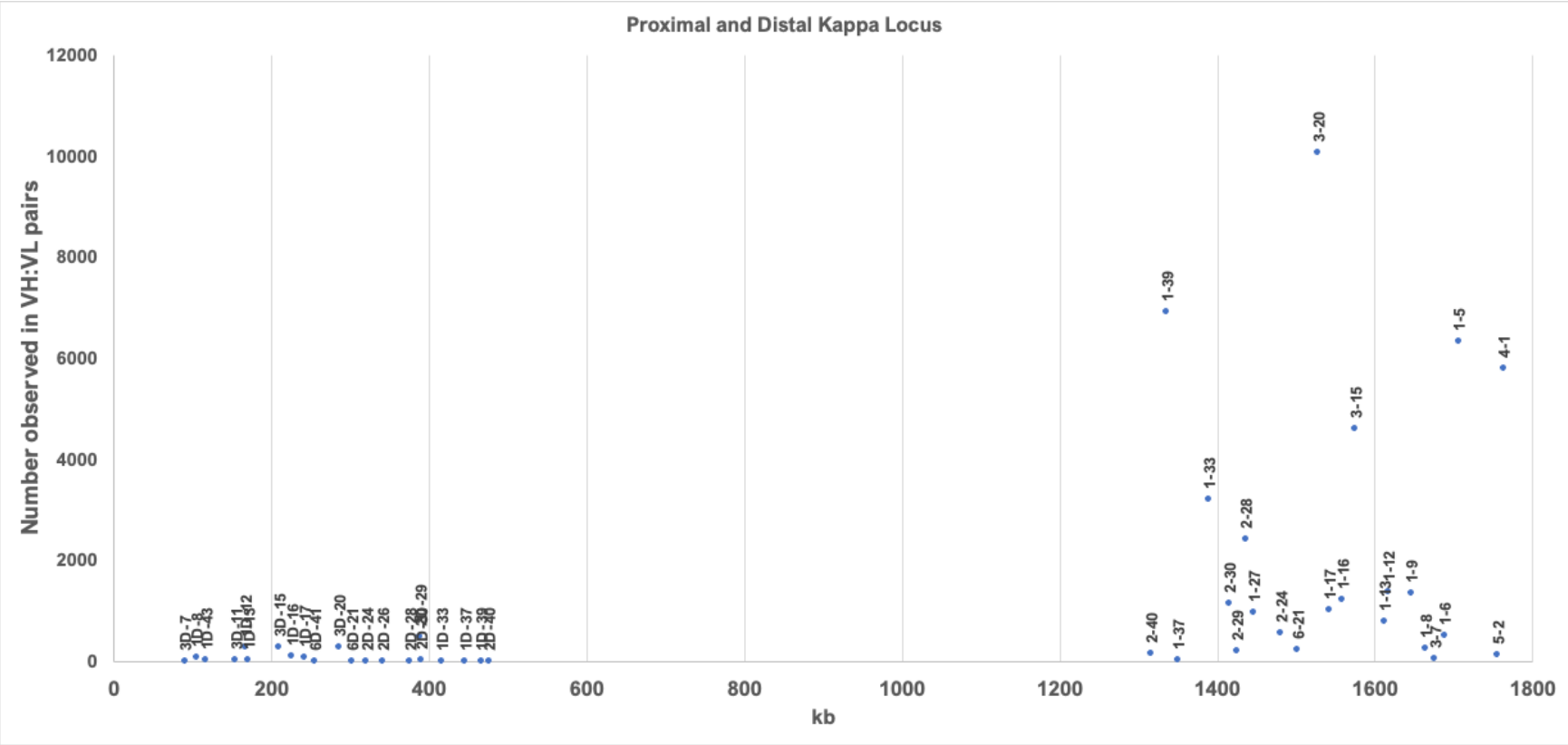

Figure 4.2. Frequency of all proximal and distal Kappa locus V-genes mapped to their genomic location. Typically, V-gene frequencies are represented in alphanumerical order, representation of frequency as a function of genomic distance, illustrates how prevalent the proximal V-genes are in contrast to the distal Vgenes.

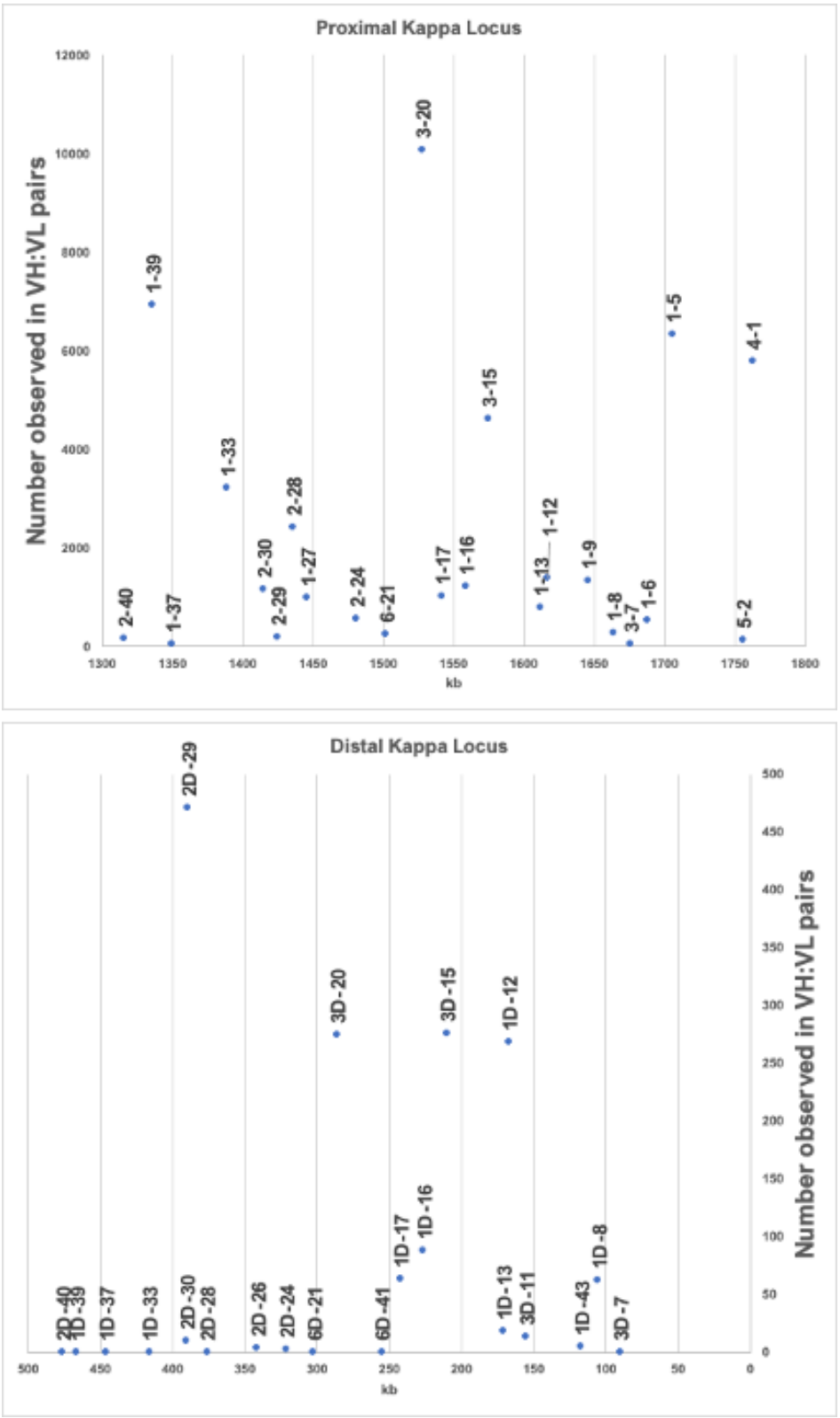

Figure 4.2. Frequency of all proximal and distal Kappa locus V-genes mapped to their genomic location

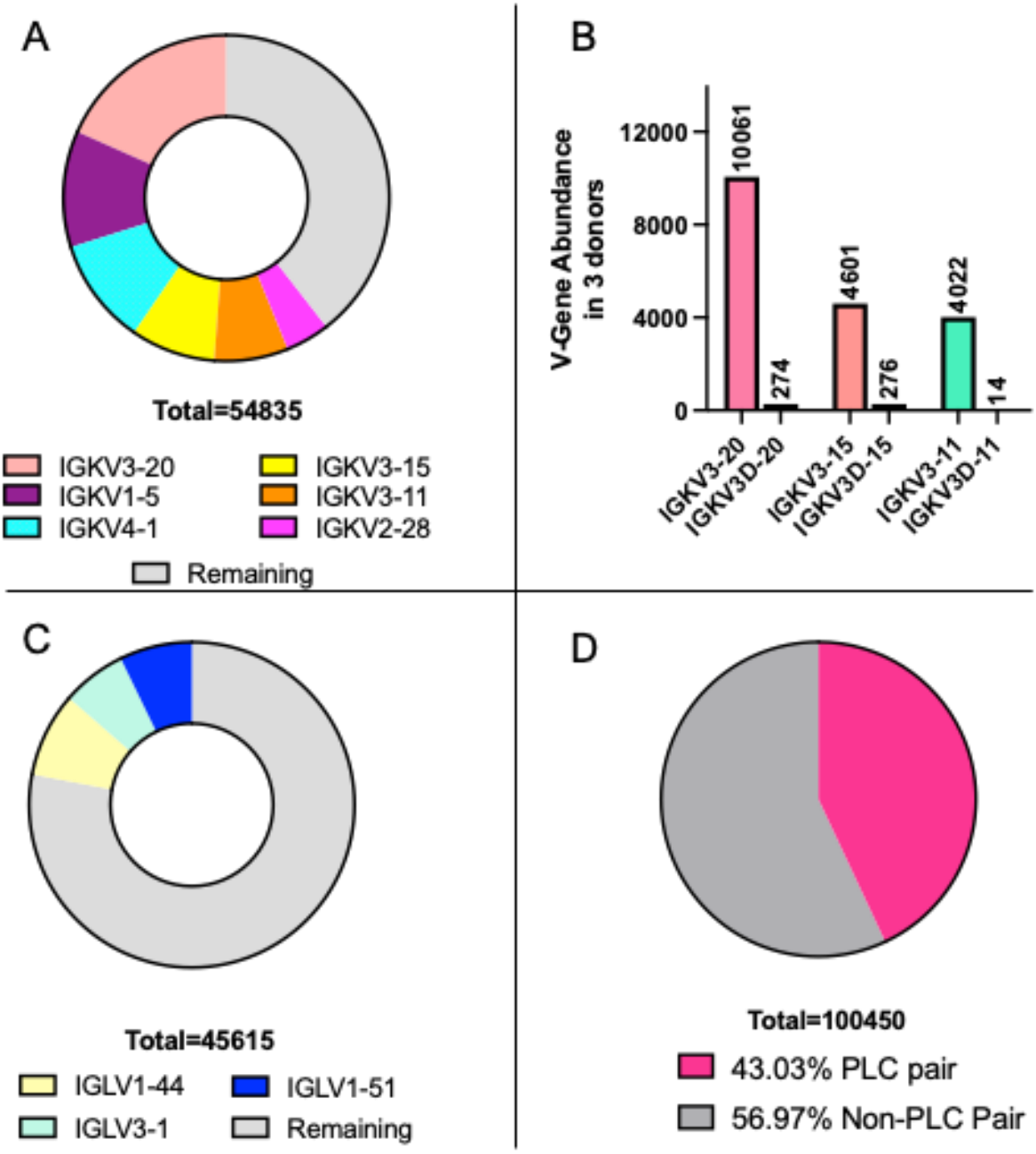

Figure 4.3. Prevalence of promiscuous light chain V-genes in donor repertoires.

Name	V-Gene	J-Gene	Light CDR1	Light CDR2	Light CDR3
PLC1	IGKV1-5	IGKJ1	QSISSW	DAS	QQYNSYSPWT
PLC2	IGKV2-28	IGKJ2	QSLLHSNGYNY	LGS	MQALQTPPYT
PLC3	IGKV3-11	IGKJ4	QSVSSY	DAS	QQRSNWPPLT
PLC4	IGKV3-15	IGKJ1	QSVSSN	GAS	QQYNNWPPWT
PLC5	IGKV3-20	IGKJ1	QSVSSSY	GAS	QQYGSSPPWT
PLC6	IGKV4-1	IGKJ2	QSVLYSSNNKNY	WAS	QQYYSTPPYT
PLC7	IGLV1-44	IGLJ3	SSNIGSNT	SNN	AAWDDSLNGPVV
PLC8	IGLV1-51	IGLJ3	SSNIGNNY	DNN	GTWDSSLSAVV
PLC9	IGLV3-1	IGLJ3	KLGDKY	QDS	QAWDSSTVV

Table 4.2 V-gene and J-gene usage of constructed PLCs.

Figure 4.4 Binding determination of VH:PLC pairs against the spike protein. A The EC50 of each mAb assayed against the ECD is represented by the color gradient. **B** Binding curves of VH 9C with each PLC. Only weak ECD binding was detected and no unspecific activity against RBD was observed.

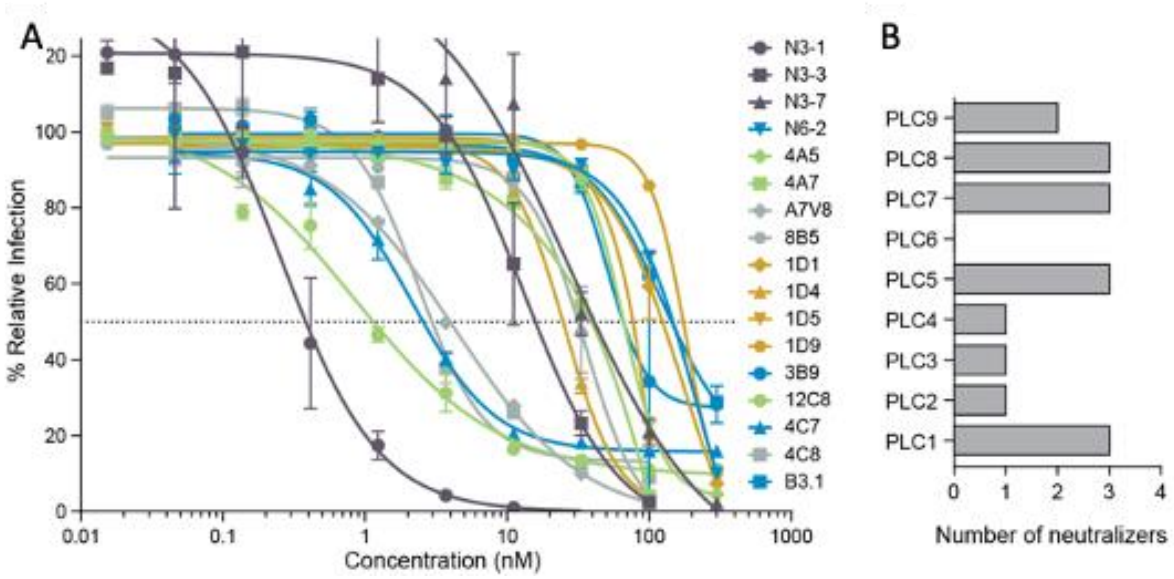

Figure 4.5 VH:PLC pairs neutralize SARS-CoV-2. A Live virus neutralization testing determined 17 neutralizing VH:PLC pairs covering a range of IC50s. **B** All PLCs produced a neutralizing mAb except for PLC6.

mAb	VH label	IC50 (nM)	VH gene	CDRH3	VL gene	CDRL3	Epitope	HC source	LC source
N3-1	N3	0.25	IGHV4-31	ARGTIYFDRSGYRRVDPFHI	IGKV1-5	QQYNSYSPWT	RBD	YSD	PLC
12C8	12C	0.84	IGHV1-24	ATGPAVRRGSWFDP	IGLV1-51	GTWDSSLSAVV	NTD	IgSeq	PLC
4C7	4C	2.15	IGHV1-24	ATAAAVRGRGTIDY	IGLV1-44	AAWDDSLNGPVV	NTD	IgSeq	PLC
4C8	4C	2.353	IGHV1-24	ATAAAVRGRGTIDY	IGLV1-51	GTWDSSLSAVV	NTD	IgSeq	PLC
A7V8	A7	4.24	IGHV1-24	ATGSPFDRTQNWFDP	IGLV1-51	GTWDSSLSAVV	NTD	IgSeq	PLC
N3-3	N3	13.1	IGHV4-31	ARGTIYFDRSGYRRVDPFHI	IGKV3-11	QQRSNWPPLT	RBD	YSD	PLC
1D4	1D	25.5	IGHV2-70	ARIPIATHLGSDY	IGKV3-15	QQYNNWPPWT	RBD	IgSeq	PLC
N3-7	N3	27.7	IGHV4-31	ARGTIYFDRSGYRRVDPFHI	IGLV1-44	AAWDDSLNGPVV	RBD	YSD	PLC
8B5	8B	34.8	IGHV1-2	ARELPPGRMVVPATYWHFDL	IGKV3-20	QQYGSSPPWT	RBD	IgSeq	PLC
3B9	3B	51.4	IGHV3-30	ARDGGGYVSY	IGLV3-1	QAWDSSTVV	NTD	IgSeq	PLC
4A5	4A	65.7	IGHV3-30-3	AKASQLFWLGQFTRDGFDI	IGKV3-20	QQYGSSPPWT	RBD	IgSeq	PLC
B3.1	B3	124	IGHV3-30	ARARGGSYYYGMDV	IGKV1-5	QQYNSYSPWT	S2	YSD	PLC
1D1	1D	173	IGHV2-70	ARIPIATHLGSDY	IGKV1-5	QQYNSYSPWT	RBD	IgSeq	PLC
1D9	1D	183	IGHV2-70	ARIPIATHLGSDY	IGLV3-1	QAWDSSTVV	RBD	IgSeq	PLC
N6-2	N6	222	IGHV3-30-3	ARPYSGSYWGYFDY	IGKV2-28	MQALQTPPYT	S2	YSD	PLC
4A7	4A	>300	IGHV3-30-3	AKASQLFWLGQFTRDGFDI	IGLV1-44	AAWDDSLNGPVV	RBD	IgSeq	PLC
1D5	1D	>300	IGHV2-70	ARIPIATHLGSDY	IGKV3-20	QQYGSSPPWT	RBD	IgSeq	PLC

Table 4.3. List of neutralizing VH:PLC antibodies. Seventeen VH:PLC pairs neutralized Wuhan-Hu-1 SARS-CoV-2 in live virus neutralization assays. Their epitopes span the S1 and S2 domains of the spike protein. Twelve pairs contained VHs identified through Ig-Seq compared to five from YSD origin. IC50s span from picomolar RBD and NTD binders, N3-1 and 12C8, to high nanomolar values for the RBD binders 4A7 and 1D5

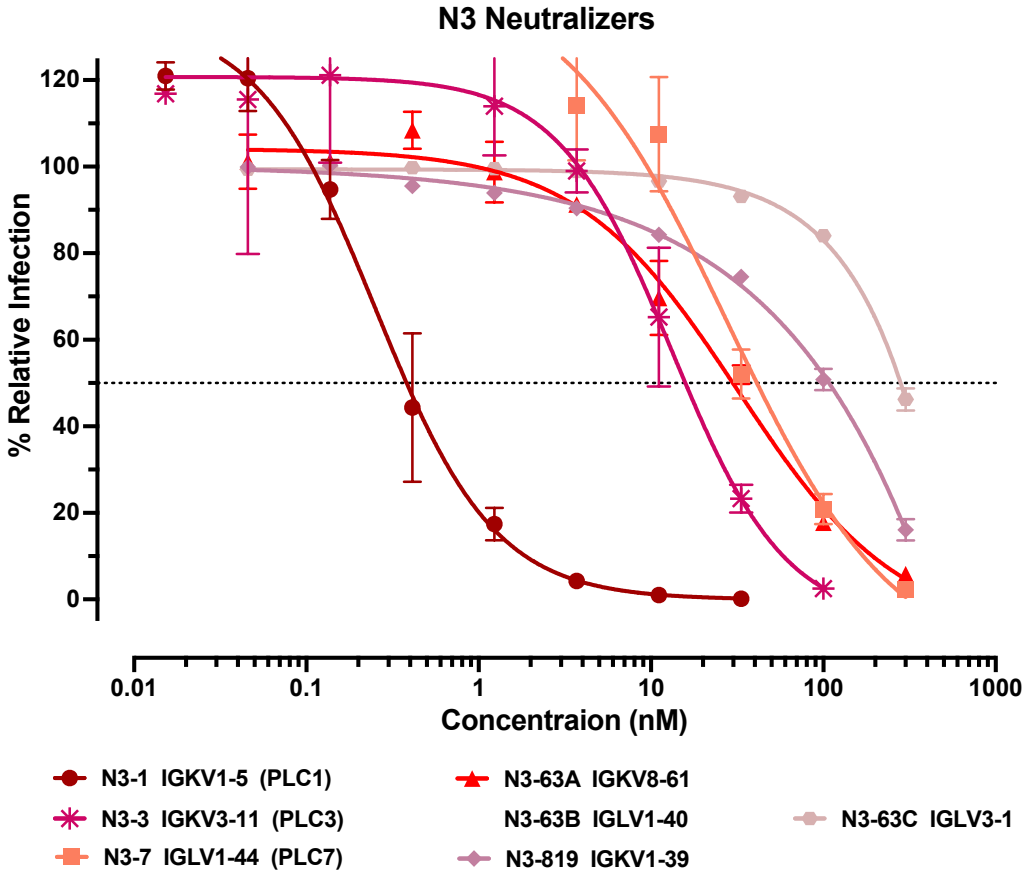

Figure 4.6. N3 forms neutralizing pairs with seven different light chains. VH N3 can pair with PLC and YSD derived light chains to produce potently neutralizing mAbs. The most potent neutralization was achieved with PLC1.

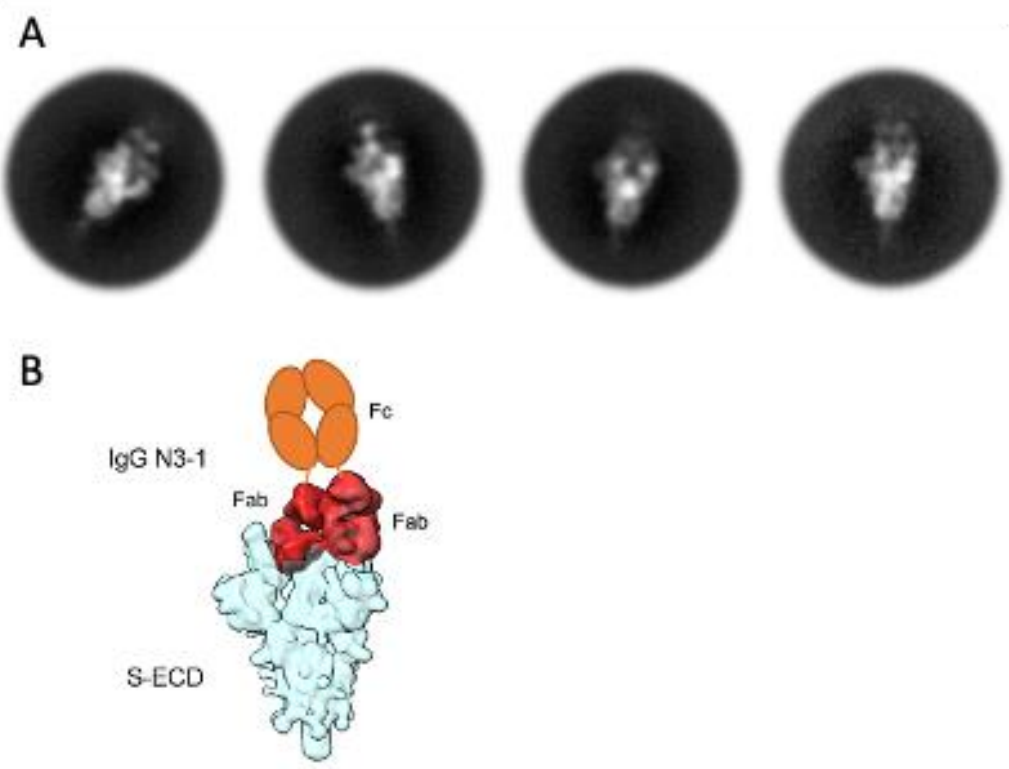

Figure 4.7. Avidity of mAb N3-1 likely achieved by a single IgG binding to a trimeric spike. A Representative 2D class averages of IgG N3-1 complexed with SARS-CoV-2 S by negative stain electron microscopy (nsEM). Although the density of Fc is not well-resolved, two clear densities of Fabs are visible. **B** A schematic model generated by Gaussian-smoothened cryoEM mapping of N3-1 bound to SARS-CoV-2 S. The Fab density is highlighted in brick red, and the spike is shown in light blue. The unobserved Fc is shown in orange.

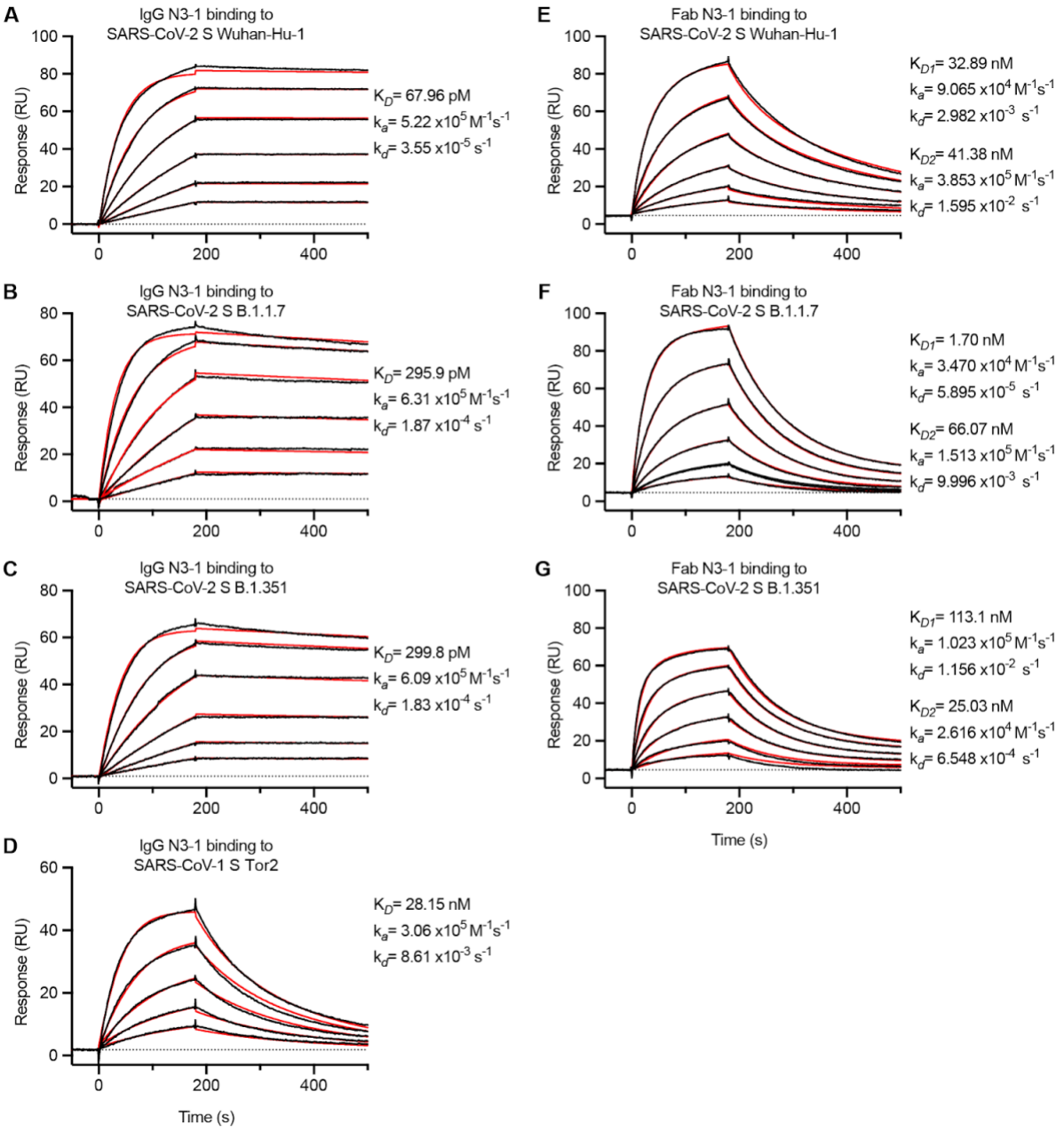

Figure 4.8. mAb N3-1 exhibits cross-reactivity and avidity to CoV spikes. A-C Binding of IgG N3-1 to SARS-CoV-2 S Wuhan-Hu-1 **A**, its variants B.1.1.7 **B** and B.1.351 **C** were assessed by surface plasmon resonance (SPR) using an NTA sensor chip. **D** Binding of IgG N3-1 to SARS-CoV-1 S was also assessed by SPR. **E-G** Binding of

Fab N3-1 to SARS-CoV-2 S Wuhan-Hu-1 **E**, B.1.1.7 **F** and B.1.351 **G**. Binding data are shown as black lines. For **A-D**, the best fit was achieved using a 1:1 binding model and shown as red lines. For **E-G**, the best fit to a heterogeneous binding model is shown as red lines.

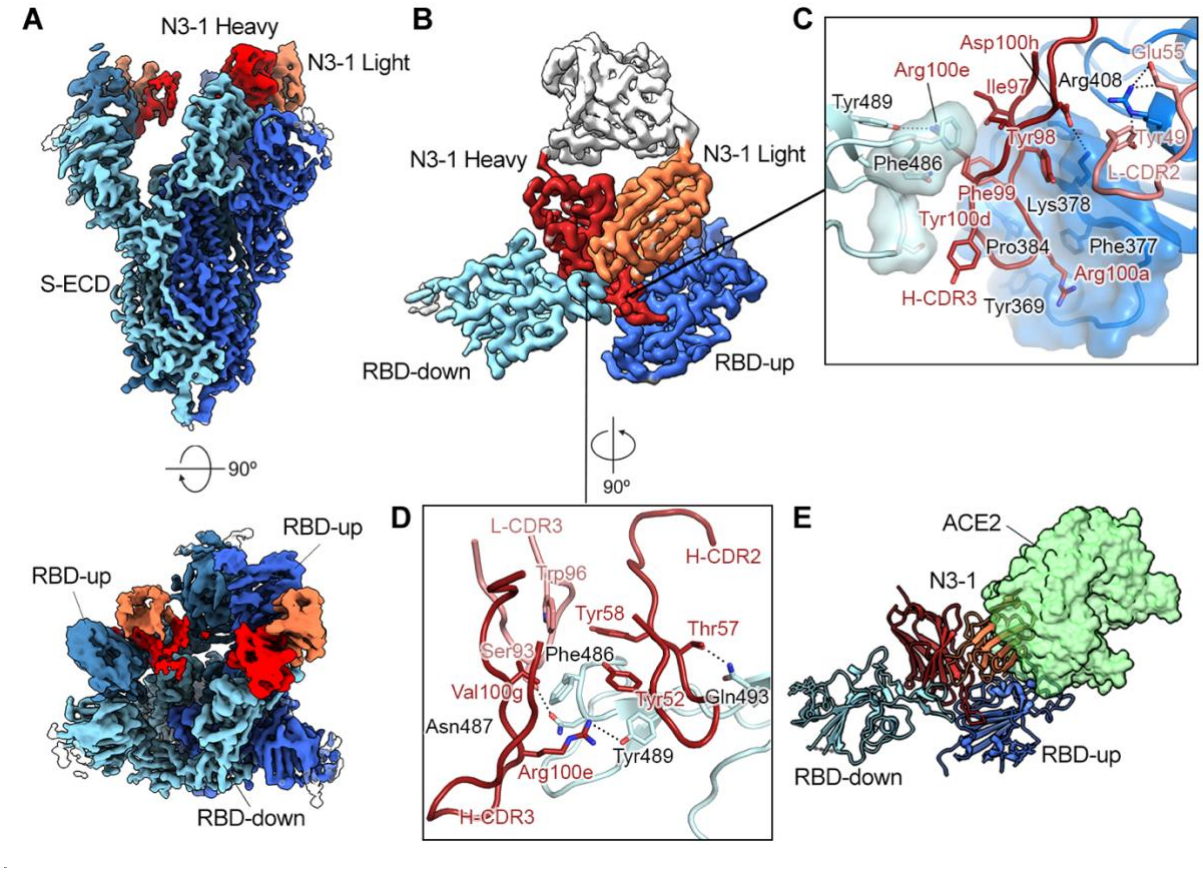

Figure 4.9. RBD-directed mAb N3-1 exhibits a unique binding mode by recognizing two distinct epitopes. A Cryo-EM structure of N3-1-bound SARS-CoV-2 spike at a global resolution of 2.8 Å. Side view and top-down views of the complex are shown in the upper and lower panel, respectively. Each protomer is depicted in steel blue, royal blue, and sky blue. The heavy chain of N3-1 is colored firebrick, and the light chain is colored coral. **B** Focused map of N3-1 bound to RBDs in the up and down conformations with five CDRs involved in the binding interface. **C** One face of CDRH3 contacts a conserved hydrophobic pocket (transparent royal blue surface) on RBD-up. The other face of CDRH3 contacts the ACE2-binding site on RBD-down. CDRH1 contacts the epitopes on RBD-up, but it is omitted for clarity. **D** The epitope on RBD-down is centered on Phe486, which fits into a hydrophobic surface formed by Trp96, Tyr58, Tyr52, Arg100e, and Val100g (clockwise). **E** Superimposed crystal structure of RBD-ACE2 complex (PDB ID: 6M0J) with N3-1 bound RBDs. The molecular surface of ACE2 is shown in transparent pale green. The light chain of N3-1 heavily clashes with ACE2. The ACE2-binding site on RBD-down is blocked by CDRH2, CDRH3, and CDRL3.

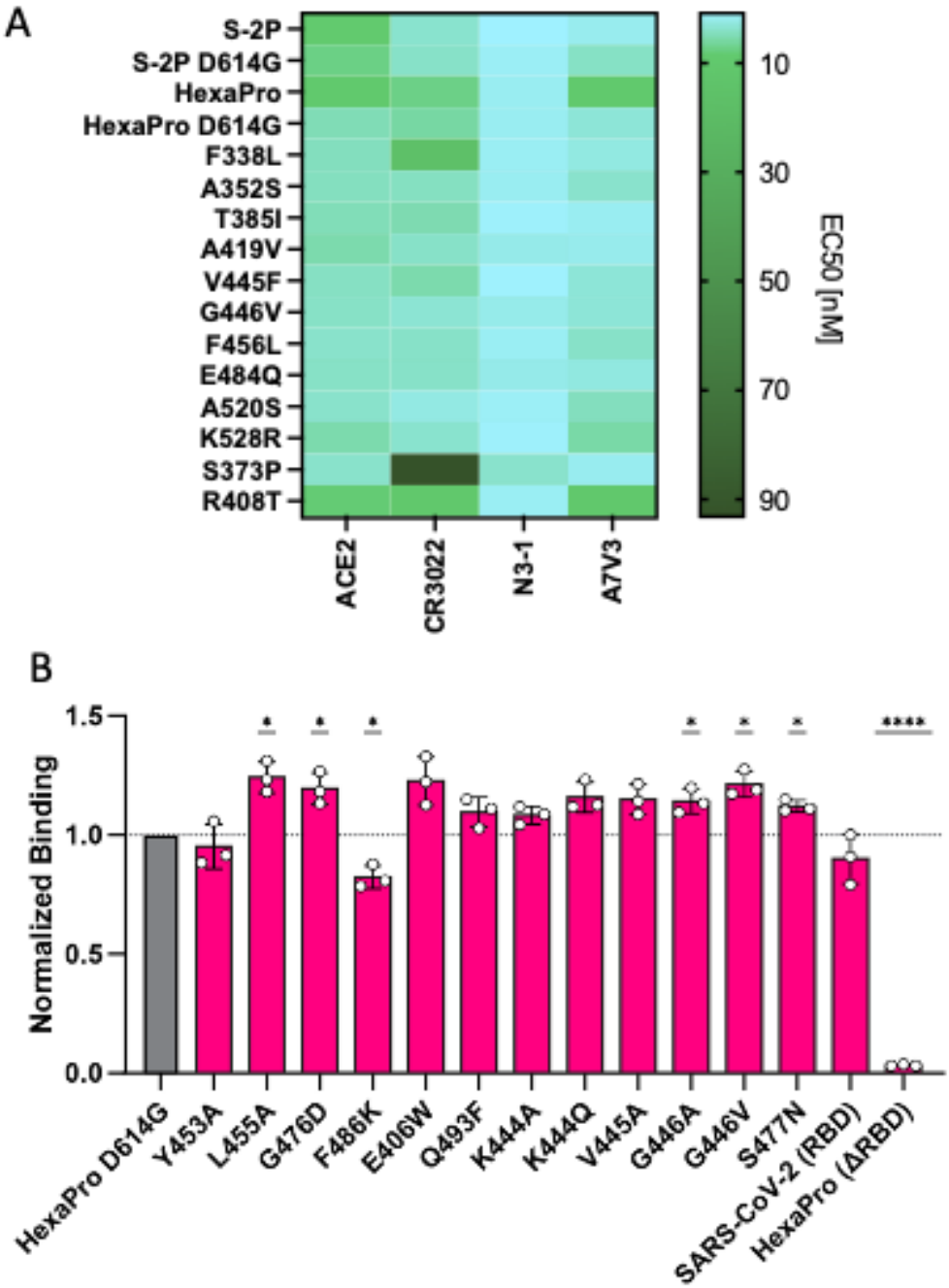

Figure 4.10 N3-1 binding to emerging SARS-CoV-2 variants. **A** N3-1 binding to variants of the spike protein circulating in Houston was tested through ELISA. **B** Mammalian spike display was used to assess N3-1 binding to circulating Regeneron antibody escape mutations. The SARS-CoV-2 RBD subunit and SARS-CoV-2 spike with a deleted RBD (ΔRBD) were included as controls. *p<0.05; ****p<0.0001; using one-sample t-test.

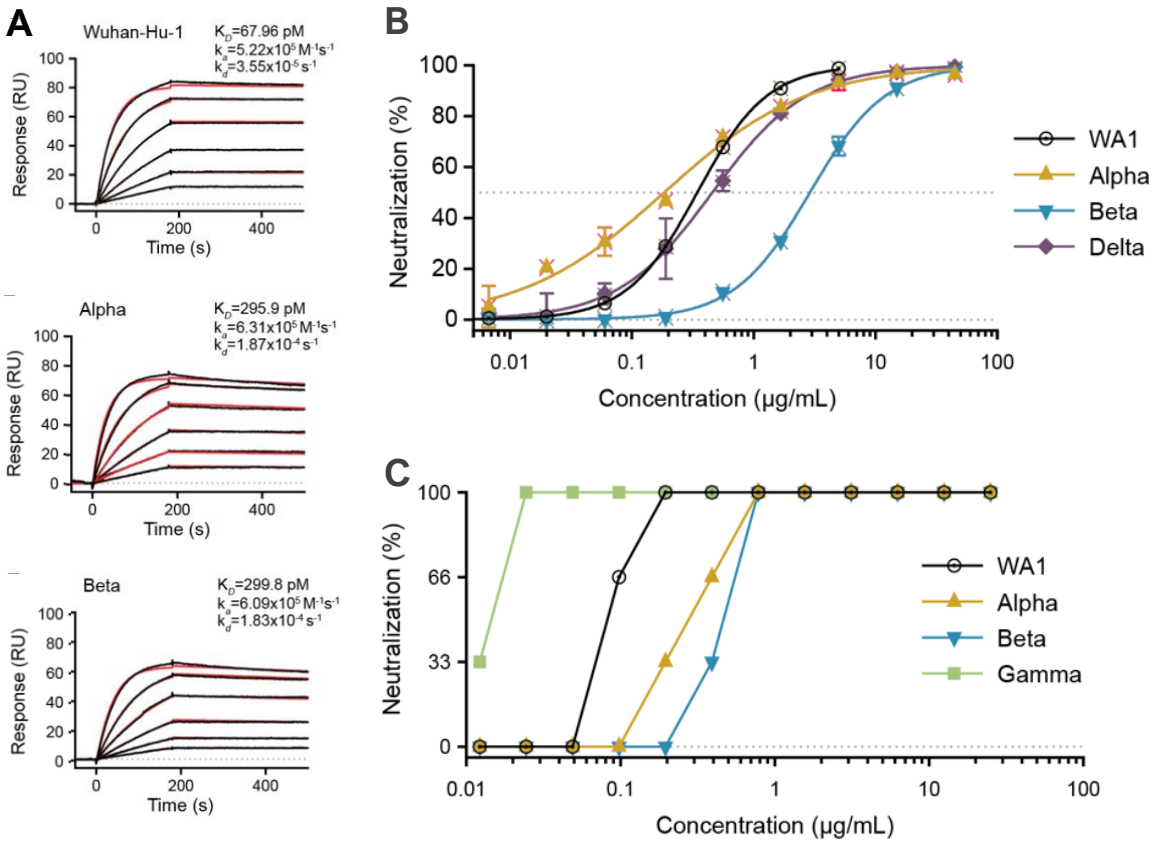

Figure 4.11 N3-1 neutralizes circulating spike variants. A SARS-CoV-2-S binding was assessed by surface plasmon resonance for Fab N3-1 binding to Wuhan-Hu-1 and IgG N3-1 binding to Wuhan-Hu-1, Alpha, and Beta. **B** and **C** Independent SARS-CoV-2 live virus neutralization assays for N3-1 IgG.

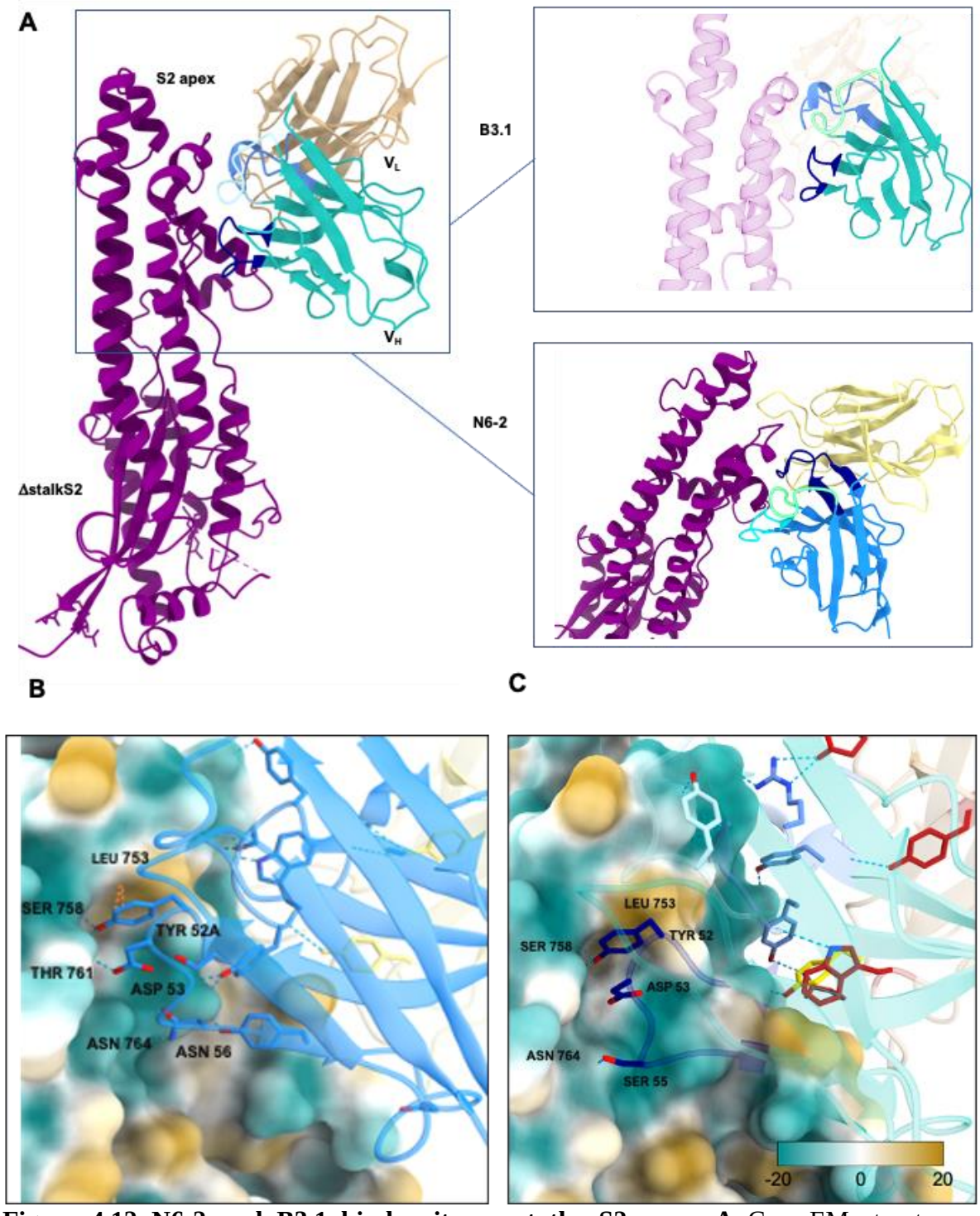

Figure 4.12 N6-2 and B3.1 bind epitopes at the S2 apex. A Cryo-EM structure determination of N6-2 and B3.1 in complex with the ΔstalkS2 identifies similar binding modes for the two antibodies at the S2 apex. **B, C** N6-2 and B.1's CDR2 is identical in

amino acid sequence and tucks into the same groove located at the S2 apex, capped by a hydrophobic ceiling from Leu753's side chain. The binding interactions of the two CDR2s differ, with N6-2 being able to engage Asp53 due to a more favorable fold that orients the residue facing the ΔstalkS2. B3.1 instead anchors its interaction through Ser55 forming hydrogen bonds with Asn764. N6-2 CDRH2 is shown in dodger blue and B3.1 CDRH2 in dark blue.

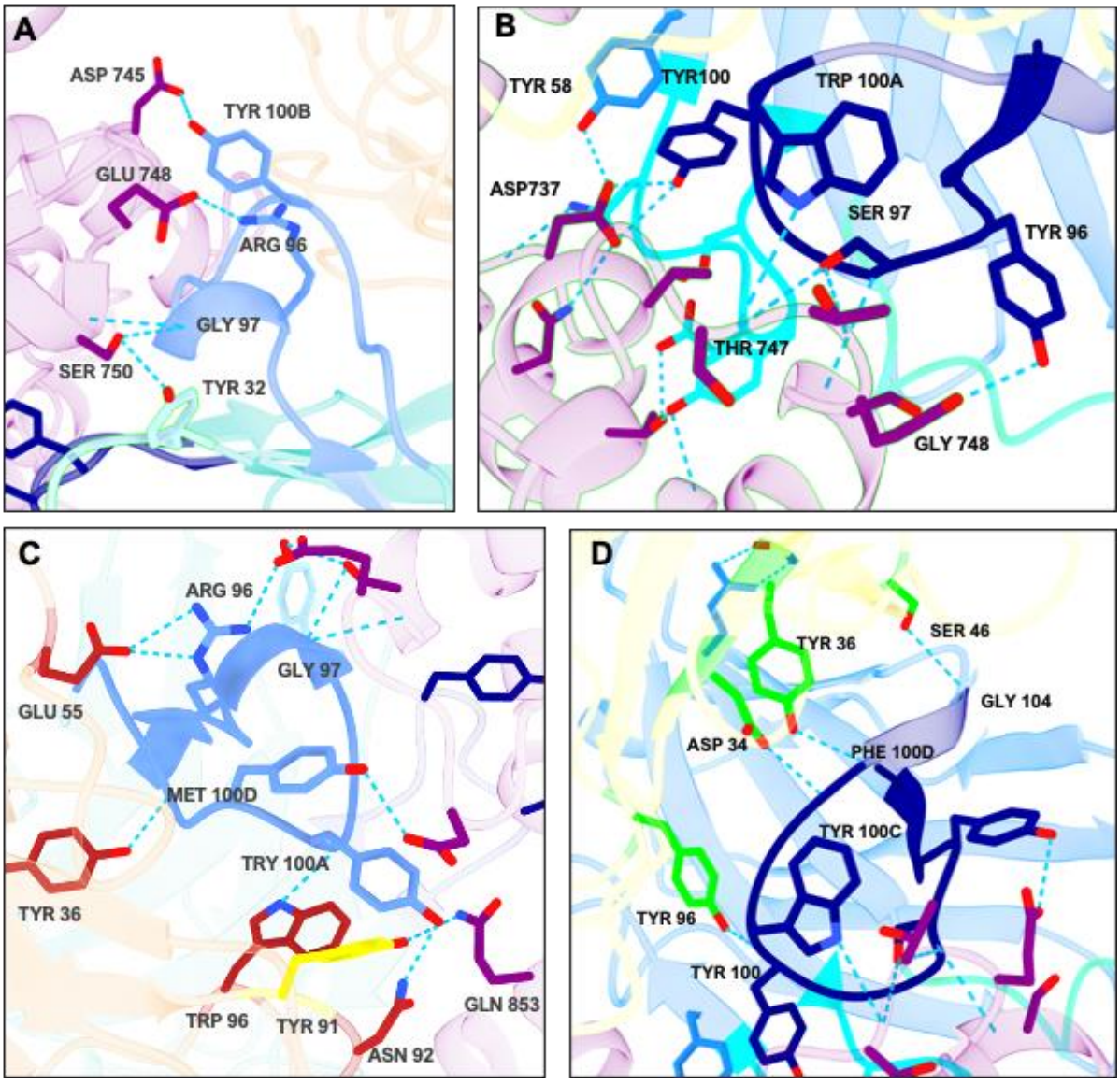

Figure 4.13 B3.1 and N6-2 CDRH3s bind different S2 epitopes. A B3.1 CDRH3 is shown in cornflower blue, CDRH1 in aquamarine and S2 in purple. **B** N6-2 CDRH3 is shown in dark blue, CDRH1 in cyan and S2 in purple. **C** Four B3.1 light chain residues in red support the CDRH3 loop through hydrogen bond interactions. Light chain Tyr91 in yellow interacts with S2's Gln853 (purple). **D** Four N6-2 light chain residues in green interact with the CDRH3 in dark blue through backbone contacts.

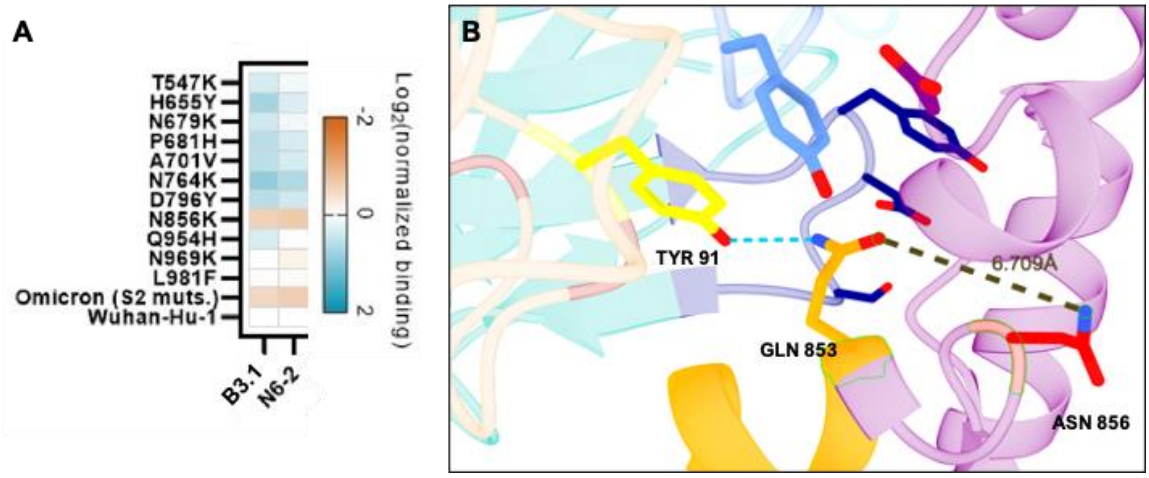

Figure 4.14 Omicron mutation N856K diminishes B3.1 and N6-2 binding. **A** Mammalian spike display was used to test the individual effects of accumulated omicron S2 mutations on B3.1 and N6-2 binding. The fold change of binding is shown as log2 as normalized to Wuhan-Hu-1 binding of each antibody. **B** The S2 residue Gln853 is located in its FPPR, colored orange, and interacts with B3.1's light chain residue Tyr91 through a hydrogen bond. Asp856 is shown in red and the distance between the amide groups was 6.709 A using ChimeraX.

CHAPTER 5: Major Findings and Future Recommendations

5.1 Combination of YSD and Ig-Seq advances antibody discovery

In this final chapter, I will examine the results of the different methods of our antibody discovery platform and delineate how these methods are most successfully applied in concert to isolate the maximum amount and greatest diversity of candidate antibodies from a discovery campaign. In total, we report 49 SARS-CoV-2 neutralizers which range in neutralization efficiency from 18 pM to >300 nM and recognize epitopes at the RBD, NTD and S2 of the Wuhan-Hu-1 S-protein (**Figure 5.1 A**). We employed YSD and Ig-Seq to source antigen-specific VHs from donor samples and identified eleven unique clonotypes from YSD and fifteen from Ig-Seq. To find productive VLs we pursued two approaches: a) screening for a suitable PLC partner, or b) expression of combinatorial Fab libraries with donor VLs to select the greatest affinity candidates via YSD. The eleven VHs discovered through YSD formed 20 neutralizing VH:VL pairs and typically paired with YSD derived VLs as well, except for B3, which only showed neutralization with PLCs. YSD-derived N3 and N6 formed neutralizing pairs with PLCs and YSD VLs.

For Ig-Seq-derived VHs, both approaches were successful and yielded 29 neutralizing VH:VL pairs of which twelve used a PLC and seventeen a YSD-derived VL. The most frequently paired YSD VLs were IGLV1-40 and IGKV1-39, followed by IGLV3-1 and IGLV1-51 which are also the V-genes of PLC8 and PLC9, respectively (**Figure 5.1 B**). A7, the VH of the NTD-binder A7V3, showed a bias towards IGLV1-51

in YSD selections, and also formed a neutralizing pair with PLC8 (IGLV1-51) but no other PLCs when screened (**Figure 5.1 C**).

This anti-SARS-CoV-2 discovery campaign was the first time PLCs were employed to identify functional pairs, which resulted in great success as they produced seventeen neutralizing mAbs, including the structurally characterized N3-1, B3.1 and N6-1. Our data demonstrates that non-cognate VH:VL pairs form neutralizing antibodies and that an antigen-specific heavy chain can partner with multiple light chains productively. In the case of IGHV1-24 clonotypes like A7 we observed a strong preference for lambda light chains, specifically IGLV1-51. Only IGHV1-24 clonotype 12C also paired with the kappa light chain IGKV1-39 (**Figure 5.2**). Other clonotypes like N3 (IGHV4-31) were less discerning in their VL partner choice and formed pairs with a variety of kappa and lambda genes. Overall, the light chain usage of our SARS-CoV-2 neutralizers was consistent with reports from other groups, which identified all PLCs as frequent partners, as well as IGKV1-39 and IGLV1-40, the most frequently YSD-derived VLs[223–227]. Our results further support observations of a convergent antibody response to SARS-CoV-2 infection, characterized by increased IGHV3 gene family usage, specifically IGHV3-30, and a class of NTD-directed neutralizers dominated by IGHV1-24 antibodies[151–153,168,181,228]. **Figure 5.2** summarizes the neutralizer VH and VL gene usage and pairing.

While the use of two different methods for heavy-chain discovery is not typical in a single discovery campaign, the complementarity of YSD and Ig-Seq resulted in the discovery of unique clonotypes from each approach that would otherwise have been missed. YSD and Ig-Seq examine spatially and temporally separate repertoires, which

increases the number and diversity of repertoire members screened. Additionally, the limitations and drawbacks inherent to each technique are mitigated by the two-pronged approach. For example, the CDRH3 of the S2 binder N6-2 contains a proline residue following the arginine residue after which trypsin cleavage for CDRH3 fragmentation typically occurs in Ig-Seq. However, the presence of the proline residue following arginine, inhibits trypsin's protease activity and leads to insufficient fragmentation of the CDRH3[229]. Instead, N6 heavy chain was discovered via YSD, which is not limited by protein sequence, although the presence of AaRI recognition sites on a nucleotide level can lead to loss of candidates. Interestingly, the additional S2 specific clonotypes N5, B3, and P4A were discovered through YSD as well, despite their CDR3H trypsin digestion not being hindered by prolines. Considering the dynamic nature of the S-protein, S2 epitopes may be more accessible to antibodies in solution during flow-cytometry-based assays compared to column-immobilized S-protein during Ig-pulldowns.

5.2 NEUTRALIZING MABS TARGET RBD, NTD AND S2 EPITOPES

We have structurally characterized the S-protein binding modes of four neutralizing mAbs: A7V3, N3-1, B3.1, and N6-2. A7V3 and N3-1 target the NTD and RBD of the S1 subunit, respectively, while N6-2 and B3.1 bind overlapping epitopes at the S2 apex. RBD binders are the most frequently observed class of neutralizers, and their mode of action typically is to directly interfere with ACE2 binding[148,149,230–232]. This is also the case for N3-1, which binds ACE2-interacting residues at the RBD and sterically blocks ACE2 approach. N3-1 differs from other RBD binders through its ability to bind

all three RBD domains with a single IgG. One of N3-1's Fabs interacts with an RBD-up and RBD-down simultaneously while the other Fab pins the third RBD in the up position. N3-1 can bind RBD in the up and down position at the same time by recognizing two discrete epitopes in the different conformational states (**Figure 4.9**). To achieve this, N3-1 employs five out of its six CDRs to bind a conserved hydrophobic pocket on the RBD-up and block the ACE2 binding site on the RBD-down. Notably, the CDRL2 of PLC1 is engaged in the hydrophobic pocket at the RBD-up while CDRL3 contacts the RBD-down, demonstrating that PLCs can contribute to antibody binding despite being a non-cognate partner. Since discovery of N3-1, other antibodies have been described that occupy two epitopes at the RBD-up and RBD-down, however, they do not bind them simultaneously but rather one at a time[233].

In contrast to N3-1's unique binding mode, A7V3 employs a common binding mode to target an antigenic supersite at the NTD of the original SARS-CoV-2 Wuhan strain. While RBD-mediated neutralization interferes directly with host recognition and therefore infection, the mechanism for NTD-directed neutralization is opaque as the NTD's function is unknown. Regardless of our lack of mechanistic understanding, the NTD is a hotspot for neutralizing antibodies, and several NTD-directed antibodies, including A7V3 (**Figure 3.9**), prophylactically protect mice from developing COVID-19 upon SARS-CoV-2 challenge[153]. NTD neutralizers overwhelmingly target a set of flexible peptide loops that lack the heavy glycosylation characteristic for most of the NTD. This antigenic supersite is vulnerable to recognition by IGHV1-24 antibodies like A7V3. The first reported NTD binder to target the site, 4A8, shares a strikingly similar

binding mode of the N3 and N5 loops with A7V3[152]. While 4A8 and A7V3 have distinct CDRH3s, their CDRH1s and CDRH2 are identical. Overlay of the two structures in complex with the S-protein (**Figure 3.8E**) reveals that the two CDRH2s form the same interactions with the N3 loop, which positions the CDRH3s to contact epitopes at the N5 loop where A7V3's and 4A8's interactions with the loop diverge due to the different CDRH3 sequences.

Interestingly, when we determined the structures of S2 binders N6-2 and B3.1 in complex with ΔstalkS2 we observed an analogous case. N6-2 and B3.1 share the IGHV3-30 V-gene while differing in their CDRH3 sequence. Examination of the binding modes shows that N6-2's and B3.1's CDRH2s occupy the same binding pocket at the S2 apex, anchored by the CDRH2s' tyrosine interacting with the hydrophobic ceiling of the pocket. While the CDRH2s' interactions with the S2 are not identical for each residue, the recognition of the same binding pocket ultimately determines the positions of the CDRH3s. Similar to the CDRH3s of A7V3 and 4A8, N6-2's and B3.1's CDRH3s bind overlapping epitopes at the same site on S2 surrounding residue Glu748.

In our screen of the effect of individual omicron mutations throughout the S2 on antibody binding, N764K in isolation improved N6-2 and B3-1 binding, most likely by strengthening the interactions at the CDRH2 binding pocket (**Figure 4.14A**), further illustrating the epitope's importance. Overall N6-2's and B3.1's binding of omicron S2 is abrogated by the non-contacting mutation N853K which may cause larger structural rearrangement leading to the occlusion of previous binding interfaces.

The similarities observed between the binding modes of antibodies of the same heavy chain V-gene may suggest an unexplored role for CDRH2s in determining antigen-specificity. It is known that specific VH-genes such as the self-reactive IGVH4-34 display biases in their antigen reactivity[234]. Considering the stochastic nature of CDRH3 sequences, VH-gene specific antigen preferences are more likely to be mediate by germline residues. Germline residue mediated antigen recognition has been demonstrated for IGHV3-30 antibodies which are disproportionally evoked by cytomegalovirus and *streptococcus pneumoniae* infection[235,236]. For cytomegalovirus recognition, CDRH2 germline residues contact a conserved peptide epitope and remain unaffected by several rounds of affinity maturation[236], indicating their functional significance. Considering these findings in the context of convergent repertoire responses to SARS-CoV-2 infection, suggests that while CDRH3 contacts are undoubtedly determinants of affinity and specificity, germline features like the CDRH2 may contribute to and bias VH-gene specific antigen recognition. A7V3 and 4A8 especially connote this hypothesis as two antibodies which were discovered by independent groups and from ethnically divers donors. Exploration of this theory through further experiments and meta-analysis of published structures may provide better insights into antibody mediated immunity in the future.

5.2 FUTURE METHOD DEVELOPMENT TO IMPROVE REPERTOIRE MINING

Currently, both YSD and Ig-Seq focus on IgG isotype antibodies, as they generally possess stronger affinities towards their antigen. However, this limits the

techniques to observation of post-class switched antibodies and misses the crucial early-response IgM antibodies. Adaptation of Ig-Seq and YSD methods for IgM discovery may provide rare insights into initial immune response against novel pathogens and enable a better understanding of repertoire dynamics during an active immune challenge. Coupling IgM and IgG repertoire analysis over the course of an immune response may grant the opportunity to trace antibody lineages from activation through class switching and affinity maturation. IgM antibodies also play significant roles in neutralization of viruses, such as Ebola[237], HIV[238] and SARS-CoV-2[239], in which case they are of interest as therapeutics.

YSD is a versatile technique that allows (immune) repertoires to be stored and repeatedly interrogated against different antigens, in competition assays, negative selections, and selection for cross-reactive binders. By creating a display strain in the laboratory strain BY4741, we hope to make YSD more accessible as an experimental tool. The reduction in induction time from two to one day shortens experimental time frames and its well characterized genetic background is compatible with common yeast biology tools such as the yeast tool kit. In future experiments we will explore whether initial improvements of Fab display in EBY100 through chaperone co-expression can be recapitulated and expanded upon in BYDis, and benefit from the OPI1-deletion mediated ER-expansion. The genomic integration of expression vectors mitigates plasmid loss in transformed yeast and significantly reduces the number of non-expressing yeast cells. Efficient display of Fabs enables more accurate characterization of antibody library members and expedited downstream adaptation of full-length antibody formats.

Yeast antibody libraries are agnostic to the repertoire source, and can be used to mine naïve, immune, and synthetic repertoires equally. One limitation of YSD can be the transformed library size. Donor derived libraries are typically extracted from PBMCs on an order of 10^6. While yeast transformation efficiency routinely reached 10^7 – 10^8 transformants, the combinatorial diversity of 10^6 VHs times 10^6 VLs far exceeds this. To mitigate the loss of library members we propose a two-step selection protocol that leverages our insight into light chain repertoire diversity. To reduce combinatorial diversity, VH-sequences can initially be paired with a limited set of light chains of germline sequence such as the PLCs to isolate antigen-specific heavy chains. In a second step these heavy chains can then be extracted and paired with the donor derived VL repertoire to find optimized pairs. To maximize discovery of antigen-specific VHs from an initial reduced light chain set, a surrogate light chain (SLC) construct can be included. The SLC's role during B-cell development is to stabilize the pre-B-cell receptor by functioning as a universal heavy chain partner[240]. During the initial selection step, VHs that do not pair productively with other light chains, could be stabilized by the SLC instead. Xu et al. successfully paired SLC constructs with heavy chains to create and select 'surrobodies' via phage display[241].

The modular design of our YSD platform enables straightforward adaptation of different expression formats such as the SLC or other Ig isotypes like IgM. It can also be amended for the display of antibodies from different species. As arguably the most relevant animal model in antibody discovery, we conducted a literature review of mice light chain repertoires and identified the most frequently occurring mice VLs (**Table**

5.1)[242,243]. While these murine PLCs have not yet been validated, PLCs for mice and other relevant species, e.g., non-human primate, may be valuable tools for antibody discovery.

In summary, the work described in this dissertation presents several advances in antibody discovery methods and illustrates how these advances enabled the rapid discovery of neutralizing SARS-CoV-2 mAbs.

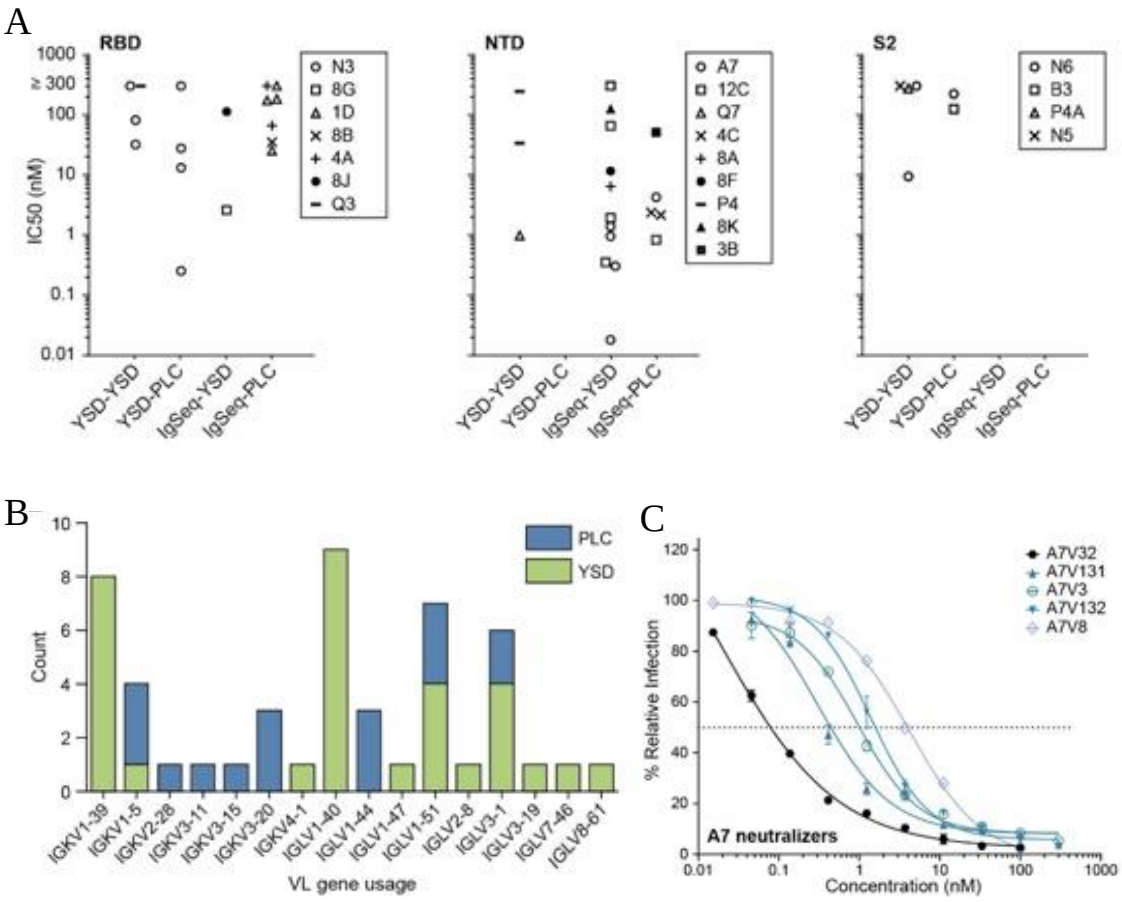

Figure 5.1 Distribution of VH:VL neutralizers by discovery method. A Neutralizing antibodies are grouped by their antigen specificity and their IC50s mapped binned by discovery method. Each unique clonotype is depicted by a symbol. Symbols occurring multiple times in the same sub-plot denote the same VH clonotype with different VLs. B plots the VL gene usage in neutralizing pairs. Blue represents PLCs and green YSD VLs. C Live virus neutralization testing of A7-derived neutralizing VH:VLs.

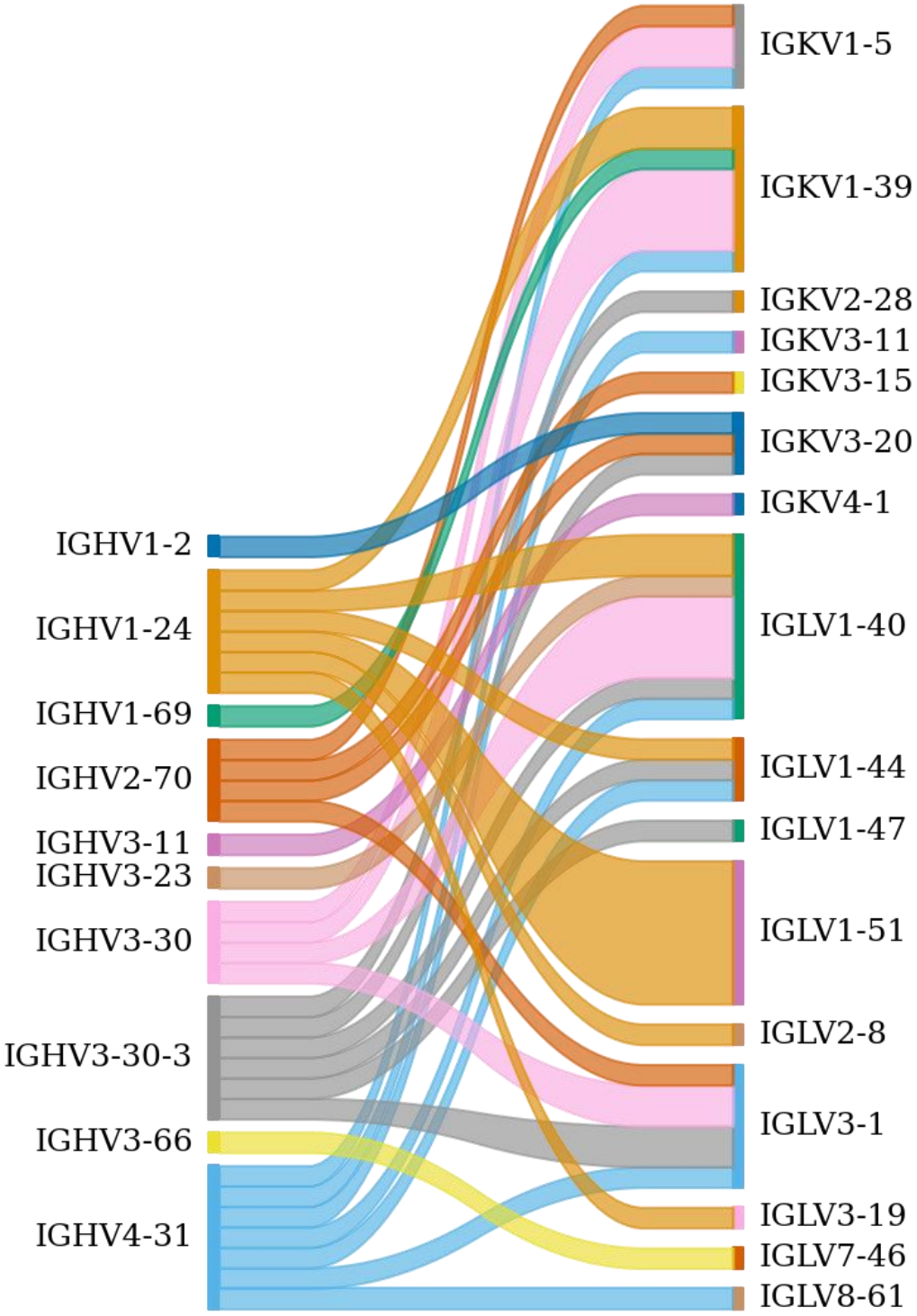

Figure 5.2 SARS-CoV-2 neutralizing antibody VH and VL gene usage and pairing. Sankey plot depicting the gene usage and VH:VL pairing of neutralizing mAbs.

V Gene	J Gene	CDR3
IGKV_1-117	J1	FQGSHVPPWT
IGKV_1-135	J1	WQGTHFPQWT
IGKV_10-96	J1	QQGNTLPRT
IGKV_12-44	J2	QHHYGTPPYT
IGKV_12-46	J1	QHFWGTPPWT
IGKV_12-89	J1	QNVLSTPPWT
IGKV_19-93	J1	LQYDNLLPWT
IGKV_6-23	J5	QQYSSYPLLT
IGKV_6-15	J5	QQYNSYPLLT
IGKV_8-30	J2	QQYYSYPPYT
IGKV_8-27	J1	HQYLSSWT
IGLV_1	J1	ALWYSNHFWV

Table 5.1 Mice PLCs. Light chain V-genes and J-genes were identified from two studies examining mouse repertoire diversity.

Bibliography

1. Behring & Kitasato. [On the development of immunity to diphtheria and tetanus in animals]. *Dtsch Med Wochenschr* **90**, 2183 (1965).
2. Ecker, D. M., Jones, S. D. & Levine, H. L. The therapeutic monoclonal antibody market. *MAbs* **7**, 9–14 (2015).
3. Barclay, A. N. Immunoglobin Superfamily. *Brenner's Encyclopedia of Genetics: Second Edition* 29–31 (2013) doi:10.1016/B978-0-12-374984-0.00772-5.
4. Galaktionov, V. G. Evolutionary Development of the Immunoglobulin Family. *Biology Bulletin* **31**, 101–111 (2004).
5. Alzari, P. M. Domains, Immunoglobulin-Type. *Encyclopedia of Immunology* 775–778 (1998) doi:10.1006/RWEI.1999.0204.
6. Schroeder, H. W., Jr & Cavacini, L. Structure and Function of Immunoglobulins. *J Allergy Clin Immunol* **125**, S41 (2010).
7. Herold, E. M. *et al.* Determinants of the assembly and function of antibody variable domains. *Scientific Reports* **7**, 12276 (2017).
8. Chiu, M. L., Goulet, D. R., Teplyakov, A. & Gilliland, G. L. Antibody Structure and Function: The Basis for Engineering Therapeutics. *Antibodies 2019, Vol. 8, Page 55* **8**, 55 (2019).
9. Charles A Janeway, J., Travers, P., Walport, M. & Shlomchik, M. J. The generation of diversity in immunoglobulins. (2001).
10. Schroeder, H. W., Cavacini, L., Birmingham, B. & Ala, B. Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology* **125**, S41–S52 (2010).
11. van der Burg, M. *et al.* Ordered recombination of immunoglobulin light chain genes occurs at the IGK locus but seems less strict at theIGL locus. *Blood* **97**, 1001–1008 (2001).
12. Agrawal, A., Eastman, Q. M. & Schatz, D. G. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**, 744–751 (1998).
13. Bassing, C. H., Swat, W. & Alt, F. W. The Mechanism and Regulation of Chromosomal V(D)J Recombination. *Cell* **109**, S45–S55 (2002).
14. Scaviner, D., Barbié, V., Ruiz, M. & Lefranc, M.-P. Protein Displays of the Human Immunoglobulin Heavy, Kappa and Lambda Variable and Joining Regions. *Experimental and Clinical Immunogenetics* **16**, 234–240 (1999).
15. Gellert, M. V(D)J Recombination: RAG Proteins, Repair Factors, and Regulation*. *http://dx.doi.org/10.1146/annurev.biochem.71.090501.150203* **71**, 101–132 (2003).
16. Rolink, A. G., Schaniel, C., Andersson, J. & Melchers, F. Selection events operating at various stages in B cell development. *Current Opinion in Immunology* **13**, 202–207 (2001).

17. Paus, D. *et al.* Antigen recognition strength regulates the choice between extrafollicular plasma cell and germinal center B cell differentiation. *J Exp Med* **203**, 1081–91 (2006).
18. Harwood, N. E. & Batista, F. D. Early Events in B Cell Activation. *Annual Review of Immunology* **28**, 185–210 (2010).
19. Heltzel, J. M. H. & Gearhart, P. J. What Targets Somatic Hypermutation to the Immunoglobulin Loci? *Viral Immunol* **33**, 277–281 (2020).
20. Maul, R. W. *et al.* Spt5 accumulation at variable genes distinguishes somatic hypermutation in germinal center B cells from ex vivo-activated cells. *Journal of Experimental Medicine* **211**, 2297–2306 (2014).
21. di Noia, J. M. & Neuberger, M. S. Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* **76**, 1–22 (2007).
22. Franklin, A. & Steele, E. J. RNA-directed DNA repair and antibody somatic hypermutation. *Trends Genet* (2021) doi:10.1016/J.TIG.2021.10.005.
23. Petermann, M. L. The Splitting of Human Gamma Globulin Antibodies by Papain and Bromelin. *J Am Chem Soc* **68**, 106–113 (2002).
24. Koenderman, L. Inside-out control of Fc-receptors. *Frontiers in Immunology* **10**, 544 (2019).
25. Übelhart, R. *et al.* Responsiveness of B cells is regulated by the hinge region of IgD. *Nature Immunology 2015 16:5* **16**, 534–543 (2015).
26. Hobeika, E., Maity, P. C. & Jumaa, H. Control of B Cell Responsiveness by Isotype and Structural Elements of the Antigen Receptor. *Trends in Immunology* **37**, 310–320 (2016).
27. Lin, J. T. & Jelinek, D. F. Immunoglobulin Structure and Function. in *Middleton's Allergy: Principles and Practice* 30–43 (Elsevier Inc, 2020).
28. Vidarsson, G., Dekkers, G. & Rispens, T. IgG Subclasses and Allotypes: From Structure to Effector Functions. *Frontiers in Immunology* **5**, (2014).
29. Aalberse, R. C., Stapel, S. O., Schuurman, J. & Rispens, T. Immunoglobulin G4: an odd antibody. *Clinical & Experimental Allergy* **39**, 469–477 (2009).
30. Adjobimey, T. & Hoerauf, A. Induction of immunoglobulin G4 in human filariasis: an indicator of immunoregulation. *http://dx.doi.org/10.1179/136485910X12786389891407* **104**, 455–464 (2013).
31. Vidarsson, G., Dekkers, G. & Rispens, T. IgG subclasses and allotypes: From structure to effector functions. *Frontiers in Immunology* **5**, 520 (2014).
32. Lu, R. M. *et al.* Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science 2020 27:1* **27**, 1–30 (2020).
33. Johnson & Johnson Reports 2020 Fourth-Quarter and Full Year Results | Johnson & Johnson. https://www.jnj.com/johnson-johnson-reports-2020-fourth-quarter-and-full-year-results.
34. 1. Humira | FiercePharma. https://www.fiercepharma.com/special-report/top-20-drugs-by-2020-sales-humira.
35. Köhler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497 (1975).

36. Strebhardt, K. & Ullrich, A. Paul Ehrlich's magic bullet concept: 100 years of progress. *Nature Reviews Cancer 2008 8:6* **8**, 473–480 (2008).
37. Pierpont, T. M., Limper, C. B. & Richards, K. L. Past, Present, and Future of Rituximab—The World's First Oncology Monoclonal Antibody Therapy. *Frontiers in Oncology* **8**, 163 (2018).
38. Maloney, D. G. *et al.* IDEC-C2B8: results of a phase I multiple-dose trial in patients with relapsed non-Hodgkin's lymphoma. *J Clin Oncol* **15**, 3266–3274 (1997).
39. McLaughlin, P. *et al.* Rituximab chimeric anti-CD20 monoclonal antibody therapy for relapsed indolent lymphoma: half of patients respond to a four-dose treatment program. *J Clin Oncol* **16**, 2825–2833 (1998).
40. Morrison, S. L., Johnson, M. J., Herzenberg, L. A. & Oi, V. T. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proc Natl Acad Sci U S A* **81**, 6851–6855 (1984).
41. Storz, U. Rituximab: How approval history is reflected by a corresponding patent filing strategy. *MAbs* **6**, 820 (2014).
42. Top 5 Highest-Selling Drugs of 2016. https://www.ajpb.com/news/top-5-highestselling-drugs-of-2016/.
43. 17. Rituxan | FiercePharma. https://www.fiercepharma.com/special-report/top-20-drugs-by-2020-sales-rituxan.
44. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature 1986 321:6069* **321**, 522–525 (1986).
45. Antibody therapeutics approved or in regulatory review in the EU or US - The Antibody Society. https://www.antibodysociety.org/resources/approved-antibodies/.
46. Weinblatt, M. E. *et al.* Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. *Arthritis Rheum* **48**, 35–45 (2003).
47. Monoclonal Antibody Therapy Market Size, Growth, Trends, 2028. https://www.fortunebusinessinsights.com/monoclonal-antibody-therapy-market-102734.
48. Ecker, D. M., Jones, S. D. & Levine, H. L. The therapeutic monoclonal antibody market. *MAbs* **7**, 9–14 (2015).
49. Trepel, F. Number and distribution of lymphocytes in man. A critical analysis. *Klin Wochenschr* **52**, 511–515 (1974).
50. Leavy, O. The birth of monoclonal antibodies. *Nature Immunology 2016 17:1* **17**, S13–S13 (2016).
51. Parray, H. A. *et al.* Hybridoma technology a versatile method for isolation of monoclonal antibodies, its applicability across species, limitations, advancement and future perspectives. *International Immunopharmacology* **85**, 106639 (2020).

52. SMITH, S. A., CROWE, J. E. & JR. Use of Human Hybridoma Technology To Isolate Human Monoclonal Antibodies. *Microbiol Spectr* **3**, AID (2015).
53. Stevens, R. H., Macy, E., Morrow, C. & Saxon, A. Characterization of a Circulating Subpopulation of Spontaneous Antitetanus Toxoid Antibody Producing B Cells Following in Vivo Booster Immunization. *The Journal of Immunology* **122**, (1979).
54. Taylor, J. J., Pape, K. A., Steach, H. R. & Jenkins, M. K. Apoptosis and antigen affinity limit effector cell differentiation of a single naïve B cell. *Science (1979)* **347**, 784 LP – 787 (2015).
55. Mahanty, S., Prigent, A. & Garraud, O. Immunogenicity of infectious pathogens and vaccine antigens. *BMC Immunology* **16**, (2015).
56. Trivedi MS *et al.* Programmed death 1 immune checkpoint inhibitors. *Clin Adv Hematol Oncol* (2015).
57. Wolchok, J. D. *et al.* Nivolumab plus ipilimumab in advanced melanoma. *N Engl J Med* **369**, 122–133 (2013).
58. Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science (1979)* **228**, 1315–1317 (1985).
59. Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. Making antibody fragments using phage display libraries. *Nature 1991 352:6336* **352**, 624–628 (1991).
60. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature 1990 348:6301* **348**, 552–554 (1990).
61. Marks, J. D. *et al.* By-passing immunization: Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222**, 581–597 (1991).
62. de Haard, H. J. *et al.* A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies. *Journal of Biological Chemistry* **274**, 18218–18230 (1999).
63. Vaughan, T. J. *et al.* Human Antibodies With Sub-Nanomolar Affinities Isolated From A Large Non-Immunized Phage Display Library. *Nature Biotechnology* **14**, 309–314 (1996).
64. Reddy, S. P., Lin, E. J., Shah, V. v. & Wu, J. J. Adalimumab. *Therapy for Severe Psoriasis* 111–126 (2016) doi:10.1016/B978-0-323-44797-3.00010-4.
65. Fitzgerald, K. A., O'Neill, L. A. J., Gearing, A. J. H. & Callard, R. E. TNFα. *The Cytokine FactsBook and Webfacts* 474–480 (2001) doi:10.1016/B978-012155142-1/50103-2.
66. Vena, G. A. & Cassano, N. Drug focus: adalimumab in the treatment of moderate to severe psoriasis. *Biologics : Targets & Therapy* **1**, 93 (2007).
67. Furst, D. E. *et al.* Adalimumab, a Fully Human Anti-Tumor Necrosis Factor-α Monoclonal Antibody, and Concomitant Standard Antirheumatic Therapy for the Treatment of Rheumatoid Arthritis: Results of STAR (Safety Trial of Adalimumab in Rheumatoid Arthritis). *Journal of Rheumatology* **30**, 2563–2571 (2003).

68. Patel, T. & Gordon, K. B. Adalimumab: Efficacy and safety in psoriasis and rheumatoid arthritis. *Dermatologic Therapy* **17**, 427–431 (2004).
69. Chan, C. E. Z., Chan, A. H. Y., Lim, A. P. C. & Hanson, B. J. Comparison of the efficiency of antibody selection from semi-synthetic scFv and non-immune Fab phage display libraries against protein targets for rapid development of diagnostic immunoassays. *J Immunol Methods* **373**, 79–88 (2011).
70. Chan, C. E. Z., Lim, A. P. C., MacAry, P. A. & Hanson, B. J. The role of phage display in therapeutic antibody discovery. *Int Immunol* **26**, 649–657 (2014).
71. Yan, X. & Xu, Z. Ribosome-display technology: applications for directed evolution of functional proteins. *Drug Discovery Today* **11**, 911–916 (2006).
72. Boder, E. T. & Wittrup, K. D. *Yeast surface display for screening combinatorial polypeptide libraries*. http://www.nature.com/naturebiotechnology (1997).
73. Zhou, C., Jacobsen, F. W., Cai, L., Chen, Q. & Shen, W. D. Development of a novel mammalian cell surface antibody display platform. *MAbs* **2**, 508–518 (2010).
74. Jakobovits, A. The long-awaited magic bullets: therapeutic human monoclonal antibodies from transgenic mice. *Expert Opin Investig Drugs* **7**, 607–14 (1998).
75. Lonberg, N. *et al.* Antigen-specific human antibodies from mice comprising four distinct genetic modifications. *Nature* **368**, 856–859 (1994).
76. Mendez, M. J. *et al.* Functional transplant of megabase human immunoglobulin loci recapitulates human antibody response in mice. *Nat Genet* **15**, 146–156 (1997).
77. Chen, J. *et al.* Immunoglobulin gene rearrangement in B cell deficient mice generated by targeted deletion of the JH locus. *Int Immunol* **5**, 647–656 (1993).
78. Mukherjee, J. *et al.* Production and characterization of protective human antibodies against Shiga toxin 1. *Infect Immun* **70**, 5896–5899 (2002).
79. Green, L. L. Antibody engineering via genetic engineering of the mouse: XenoMouse strains are a vehicle for the facile generation of therapeutic human monoclonal antibodies. *J Immunol Methods* **231**, 11–23 (1999).
80. Green, L. L. *et al.* Antigen-specific human monoclonal antibodies from mice engineered with human Ig heavy and light chain YACs. *Nat Genet* **7**, 13–21 (1994).
81. Gibson, T. B., Ranganathan, A. & Grothey, A. Randomized phase III trial results of panitumumab, a fully human anti-epidermal growth factor receptor monoclonal antibody, in metastatic colorectal cancer. *Clin Colorectal Cancer* **6**, 29–31 (2006).
82. Berardi, R. *et al.* Panitumumab: the evidence for its use in the treatment of metastatic colorectal cancer. *Core Evid* **5**, 61–76 (2010).
83. Simon, E. G., Samuel, S., Ghosh, S. & Moran, G. W. Ustekinumab: a novel therapeutic option in Crohn's disease. *Expert Opinion on Biological Therapy* **16**, 1065–1074 (2016).
84. Fiorino, G. *et al.* Positioning ustekinumab in moderate-to-severe ulcerative colitis: new kid on the block. *Expert Opin Biol Ther* **20**, 421–427 (2020).

85. Ghosh, S. *et al.* Ustekinumab Safety in Psoriasis, Psoriatic Arthritis, and Crohn's Disease: An Integrated Analysis of Phase II/III Clinical Development Programs. *Drug Saf* **42**, (2019).
86. Lee, E. C. *et al.* Complete humanization of the mouse immunoglobulin loci enables efficient therapeutic antibody discovery. *Nat Biotechnol* **32**, 356–363 (2014).
87. Osborn, M. J. *et al.* High-affinity IgG antibodies develop naturally in Ig-knockout rats carrying germline human IgH/Igκ/Igλ loci bearing the rat CH region. *J Immunol* **190**, 1481–1490 (2013).
88. Kwon, Y. J. & Forthal, D. COVID-19: An unprecedented challenge and an opportunity for change. *Advanced Drug Delivery Reviews* **171**, 48–49 (2021).
89. Hansen, J. *et al.* Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science (1979)* **369**, 1010–1014 (2020).
90. Baum, A. *et al.* Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* (2020) doi:10.1126/science.abd0831.
91. Weinreich, D. M. *et al.* REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19. *New England Journal of Medicine* **384**, 238–251 (2021).
92. Coronavirus (COVID-19) Update: FDA Authorizes Monoclonal Antibodies for Treatment of COVID-19 | FDA. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-monoclonal-antibodies-treatment-covid-19.
93. Nagano, K. & Tsutsumi, Y. Phage Display Technology as a Powerful Platform for Antibody Drug Discovery. *Viruses 2021, Vol. 13, Page 178* **13**, 178 (2021).
94. Feige, M. J., Hendershot, L. M. & Buchner, J. How antibodies fold. *Trends in Biochemical Sciences* (2010) doi:10.1016/j.tibs.2009.11.005.
95. Patil, C. & Walter, P. Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. *Current Opinion in Cell Biology* **13**, 349–355 (2001).
96. Boder, E. T. & Wittrup, K. D. Optimal Screening of Surface-Displayed Polypeptide Libraries. *Biotechnology Progress* **14**, 55–62 (1998).
97. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proceedings of the National Academy of Sciences* (2000) doi:10.1073/pnas.170297297.
98. Daugherty, P. S., Iverson, B. L. & Georgiou, G. *Flow cytometric screening of cell-based libraries*. *Journal of Immunological Methods* vol. 243 (2000).
99. Röthlisberger, D., Honegger, A. & Plückthun, A. Domain Interactions in the Fab Fragment: A Comparative Evaluation of the Single-chain Fv and Fab Format Engineered with Variable Domains of Different Stability. *Journal of Molecular Biology* **347**, 773–789 (2005).

100. Quintero-Hernández, V. *et al.* The change of the scFv into the Fab format improves the stability and in vivo toxin neutralization capacity of recombinant antibodies. *Molecular Immunology* **44**, 1307–1315 (2007).
101. Sivelle, C. *et al.* Fab is the most efficient format to express functional antibodies by yeast surface display. *MAbs* **10**, 720–729 (2018).
102. Yang, Z. *et al.* Affinity maturation of an TpoR targeting antibody in full-length IgG form for enhanced agonist activity. *Protein Engineering, Design and Selection* **31**, 233–241 (2018).
103. Nguyen, A. W., Le, K. C. & Maynard, J. A. Identification of high affinity HER2 binding antibodies using CHO Fab surface display. *Protein Engineering, Design and Selection* **31**, 91–101 (2018).
104. Ho, M. & Pastan, I. Mammalian cell display for antibody engineering. *Methods Mol Biol* **525**, 337–52, xiv (2009).
105. Shusta, E. v., Raines, R. T., Plückthun, A. & Wittrup, K. D. Increasing the secretory capacity of Saccharomyces cerevisiae for production of single-chain antibody fragments. *Nature Biotechnology* **16**, 773–777 (1998).
106. Feige, M. J. *et al.* An Unfolded CH1 Domain Controls the Assembly and Secretion of IgG Antibodies. *Molecular Cell* **34**, 569–579 (2009).
107. Lee, Y. K., Brewer, J. W., Hellman, R. & Hendershot, L. M. BiP and immunoglobulin light chain cooperate to control the folding of heavy chain and ensure the fidelity of immunoglobulin assembly. *Mol Biol Cell* **10**, 2209–19 (1999).
108. Roth, R. A. & Pierce, S. B. In Vivo Cross-Linking of Protein Disulfide Isomerase to Immunoglobulins. *Biochemistry* **26**, 4179–4182 (1987).
109. Mayer, M., Kies, U., Kammermeier, R. & Buchner, J. BiP and PDI Cooperate in the Oxidative Folding of Antibodies in vitro. *Journal of Biological Chemistry* **275**, 29421–29425 (2000).
110. Jansen, G. *et al.* An interaction map of endoplasmic reticulum chaperones and foldases. *Mol Cell Proteomics* **11**, 710–23 (2012).
111. Zhu, H. *et al.* Ufbp1 promotes plasma cell development and ER expansion by modulating distinct branches of UPR. *Nature Communications 2019 10:1* **10**, 1–15 (2019).
112. Mori, K., Ogawa, N., Kawahara, T., Yanagi, H. & Yura, T. mRNA splicing-mediated C-terminal replacement of transcription factor Hac1p is required for efficient activation of the unfolded protein response. *Proc Natl Acad Sci U S A* **97**, 4660–5 (2000).
113. Mori, K., Ogawa, N., Kawahara, T., Yanagi, H. & Yura, T. Palindrome with spacer of one nucleotide is characteristic of the cis-acting unfolded protein response element in Saccharomyces cerevisiae. *J Biol Chem* **273**, 9912–20 (1998).
114. Chawla, A., Chakrabarti, S., Ghosh, G. & Niwa, M. Attenuation of yeast UPR is essential for survival and is mediated by IRE1 kinase. *The Journal of Cell Biology* **193**, 41–50 (2011).

115. Schuck, S., Prinz, W. A., Thorn, K. S., Voss, C. & Walter, P. Membrane expansion alleviates endoplasmic reticulum stress independently of the unfolded protein response. *The Journal of Cell Biology* **187**, 525–536 (2009).
116. de Ruijter, J. C., Koskela, E. V. & Frey, A. D. Enhancing antibody folding and secretion by tailoring the Saccharomyces cerevisiae endoplasmic reticulum. *Microbial Cell Factories* **15**, 87 (2016).
117. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology* (2015) doi:10.1021/sb500366v.
118. Engler, C. & Marillonnet, S. Golden Gate cloning. *Methods Mol Biol* **1116**, 119–131 (2014).
119. Triller, G. *et al.* Natural Parasite Exposure Induces Protective Human Anti-Malarial Antibodies. *Immunity* **47**, 1197-1209.e10 (2017).
120. Collins, K. A., Snaith, R., Cottingham, M. G., Gilbert, S. C. & Hill, A. V. S. Enhancing protective immunity to malaria with a highly immunogenic virus-like particle vaccine. *Sci Rep* **7**, 46621 (2017).
121. de Ruijter, J. C., Koskela, E. v. & Frey, A. D. Enhancing antibody folding and secretion by tailoring the Saccharomyces cerevisiae endoplasmic reticulum. *Microbial Cell Factories* **15**, (2016).
122. Monteilhet, C., Perrin, A., Thierry, A., Colleaux, L. & Dujon, B. Purification and characterization of the in vitro activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron. *Nucleic Acids Research* **18**, 1407–1413 (1990).
123. Jones, E. W. Tackling the Protease Problem in Saccharomyces cerevisiae. *Methods in Enzymology* **194**, 428–453 (1991).
124. Sprague, G. F. & Herskowitz, I. Control of yeast cell type by the mating type locus. I. Identification and control of expression of the a-specific gene BAR1. *J Mol Biol* **153**, 305–321 (1981).
125. Akhmetov, A. *et al.* Single-step Precision Genome Editing in Yeast Using CRISPR-Cas9. *Bio Protoc* **8**, (2018).
126. Haas, I. G. & Wabl, M. Immunoglobulin heavy chain binding protein. *Nature* (1983) doi:10.1038/306387a0.
127. Lin, H. Y. *et al.* The 170-kDa glucose-regulated stress protein is an endoplasmic reticulum protein that binds immunoglobulin. *Mol Biol Cell* **4**, 1109–19 (1993).
128. Meunier, L. A Subset of Chaperones and Folding Enzymes Form Multiprotein Complexes in Endoplasmic Reticulum to Bind Nascent Proteins. *Molecular Biology of the Cell* **13**, 4456–4469 (2002).
129. Mayer, M., Frey, S., Koivunen, P., Myllyharju, J. & Buchner, J. Influence of the Oxidoreductase ERp57 on the Folding of an Antibody Fab Fragment. *Journal of Molecular Biology* **341**, 1077–1084 (2004).
130. H5N1 avian influenza: Timeline of major events. (2011).
131. Influenza A (H1N1) outbreak . https://www.who.int/emergencies/situations/influenza-a-(h1n1)-outbreak.

132. Zika virus. https://www.who.int/news-room/fact-sheets/detail/zika-virus.
133. Ebola virus disease. https://www.who.int/health-topics/ebola/#tab=tab_1.
134. History of 1918 Flu Pandemic | Pandemic Influenza (Flu) | CDC. https://www.cdc.gov/flu/pandemic-resources/1918-commemoration/1918-pandemic-history.htm.
135. Holmes, E. C. *et al.* The origins of SARS-CoV-2: A critical review. *Cell* **184**, 4848–4856 (2021).
136. Lytras, S., Xia, W., Hughes, J., Jiang, X. & Robertson, D. L. The animal origin of SARS-CoV-2. *Science (1979)* **373**, 968–970 (2021).
137. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
138. Tian, X. *et al.* Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerging Microbes and Infections* **9**, 382–385 (2020).
139. Tea, F. *et al.* SARS-CoV-2 neutralizing antibodies: Longevity, breadth, and evasion by emerging viral variants. *PLOS Medicine* **18**, e1003656 (2021).
140. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).
141. Tai, W. *et al.* Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular & Molecular Immunology 2020 17:6* **17**, 613–620 (2020).
142. Yan, R. *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (1979)* **367**, 1444–1448 (2020).
143. Wrapp, D. *et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation*. *science.sciencemag.org* http://science.sciencemag.org/ (2019).
144. Klinakis, A., Cournia, Z. & Rampias, T. N-terminal domain mutations of the spike protein are structurally implicated in epitope recognition in emerging SARS-CoV-2 strains. *Computational and Structural Biotechnology Journal* **19**, 5556–5567 (2021).
145. Koppisetti, R. K., Fulcher, Y. G. & van Doren, S. R. Fusion Peptide of SARS-CoV-2 Spike Rearranges into a Wedge Inserted in Bilayered Micelles. *J Am Chem Soc* **143**, 13205–13211 (2021).
146. Berger, I. & Schaffitzel, C. The SARS-CoV-2 spike protein: balancing stability and infectivity. *Cell Research 2020 30:12* **30**, 1059–1060 (2020).
147. Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, (2020).
148. Yi, C. *et al.* Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell Mol Immunol* **17**, 621–630 (2020).
149. Shi, R. *et al.* A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature 2020 584:7819* **584**, 120–124 (2020).

150. Yuan, M. *et al.* A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science (1979)* **368**, 630–633 (2020).
151. McCallum, M. *et al.* N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**, 2332-2347.e16 (2021).
152. Chi, X. *et al.* A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science (1979)* **369**, 650–655 (2020).
153. Voss, W. N. *et al.* Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes. *Science* **372**, 1108–1112 (2021).
154. Baum, A. *et al.* Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science (1979)* **369**, 1014–1018 (2020).
155. Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827.e19 (2020).
156. Zhang, L. *et al.* The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv* (2020) doi:10.1101/2020.06.12.148726.
157. Wang, P. *et al.* Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature 2021 593:7857* **593**, 130–135 (2021).
158. Shen, X. *et al.* SARS-CoV-2 variant B.1.1.7 is susceptible to neutralizing antibodies elicited by ancestral spike vaccines. *Cell Host and Microbe* **29**, 529-539.e3 (2021).
159. Ferrareze, P. A. G. *et al.* E484K as an innovative phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil. *Infection, Genetics and Evolution* **93**, 104941 (2021).
160. Long, S. W. *et al.* Molecular architecture of early dissemination and massive second wave of the SARS-CoV-2 virus in a major metropolitan area. *mBio* (2020) doi:10.1128/mBio.02707-20.
161. Mlcochova, P. *et al.* SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* **599**, (2021).
162. Saville, J. W. *et al.* Structural and biochemical rationale for enhanced spike protein fitness in delta and kappa SARS-CoV-2 variants. *Nature Communications 2022 13:1* **13**, 1–10 (2022).
163. Tian, D., Sun, Y., Zhou, J. & Ye, Q. The Global Epidemic of the SARS-CoV-2 Delta Variant, Key Spike Mutations and Immune Escape. *Frontiers in Immunology* **12**, 5001 (2021).
164. Lavinder, J. J., Horton, A. P., Georgiou, G. & Ippolito, G. C. Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires. *Current Opinion in Chemical Biology* (2015) doi:10.1016/j.cbpa.2014.11.007.
165. Lavinder, J. J. *et al.* Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proceedings of the National Academy of Sciences* (2014) doi:10.1073/pnas.1317793111.

166. Lee, J. *et al.* Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nature Medicine* (2016) doi:10.1038/nm.4224.
167. Zost, S. J. *et al.* Potently neutralizing and protective human antibodies against SARS-CoV-2. *Nature* **584**, 443–449 (2020).
168. Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature 2020 584:7821* **584**, 437–442 (2020).
169. Voss, W. N. *et al.* Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes in COVID-19 convalescent plasma. *bioRxiv* (2020).
170. Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C.-M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Engineering, Design and Selection* **23**, 155–159 (2010).
171. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods 2019 16:11* **16**, 1146–1152 (2019).
172. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods 2017 14:3* **14**, 290–296 (2017).
173. Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *bioRxiv* 2020.06.12.148296 (2020) doi:10.1101/2020.06.12.148296.
174. Chi, X. *et al.* A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science (1979)* **369**, 650–655 (2020).
175. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14–25 (2018).
176. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).
177. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948–1954 (2002).
178. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).
179. Hou, Y. J. *et al.* SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell* **182**, 429-446.e14 (2020).
180. Leist, S. R. *et al.* A Mouse-Adapted SARS-CoV-2 Induces Acute Lung Injury and Mortality in Standard Laboratory Mice. *Cell* **183**, 1070-1085.e12 (2020).
181. Cerutti, G. *et al.* Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host & Microbe* **29**, 819-833.e7 (2021).
182. Klinakis, A., Cournia, Z. & Rampias, T. N-terminal domain mutations of the spike protein are structurally implicated in epitope recognition in emerging SARS-CoV-2 strains. *Computational and Structural Biotechnology Journal* **19**, 5556 (2021).

183. Martin, D. P. *et al.* Selection analysis identifies unusual clustered mutational changes in Omicron lineage BA.1 that likely impact Spike function. *bioRxiv* 2022.01.14.476382 (2022) doi:10.1101/2022.01.14.476382.
184. Viana, R. *et al.* Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* **603**, (2022).
185. Cele, S. *et al.* SARS-CoV-2 Omicron has extensive but incomplete escape of Pfizer BNT162b2 elicited neutralization and requires ACE2 for infection. *medRxiv* **11**, 2021.12.08.21267417 (2021).
186. Starr, T. N. *et al.* SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* **597**, 97–102 (2021).
187. Tortorici, M. A. *et al.* Broad sarbecovirus neutralization by a human monoclonal antibody. *Nature* **597**, 103–108 (2021).
188. Park, Y. J. *et al.* Antibody-mediated broad sarbecovirus neutralization through ACE2 molecular mimicry. *Science (1979)* **375**, (2022).
189. VanBlargan, L. A. *et al.* An infectious SARS-CoV-2 B.1.1.529 Omicron virus escapes neutralization by therapeutic monoclonal antibodies. *Nature Medicine 2022 28:3* **28**, 490–495 (2022).
190. Cameroni, E. *et al.* Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature 2021 602:7898* **602**, 664–670 (2021).
191. Tan, C.-W. *et al.* Pan-Sarbecovirus Neutralizing Antibodies in BNT162b2-Immunized SARS-CoV-1 Survivors. *New England Journal of Medicine* **385**, 1401–1406 (2021).
192. Starr, T. N. *et al.* SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* **597**, 97–102 (2021).
193. Kwon, D. This 'super antibody' for COVID fights off multiple coronaviruses. *Nature* (2021) doi:10.1038/D41586-021-01917-9.
194. Cox, J. P. L., Tomlinson, I. M. & Winter, G. A directory of human germ-line Vχ segments reveals a strong bias in their usage. *European Journal of Immunology* **24**, 827–836 (1994).
195. Weber, J. C., Blaison, G., Martin, T., Knapp, A. M. & Pasquali, J. L. Evidence that the V kappa III gene usage is nonstochastic in both adult and newborn peripheral B cells and that peripheral CD5+ adult B cells are oligoclonal. *The Journal of Clinical Investigation* **93**, 2093–2105 (1994).
196. Dono, M., Cerruti, G. & Zupo, S. The CD5+ B-cell. *The International Journal of Biochemistry & Cell Biology* **36**, 2105–2111 (2004).
197. Foster, S. J., Brezinschek, H. P., Brezinschek, R. I. & Lipsky, P. E. Molecular mechanisms and selective influences that shape the kappa gene repertoire of IgM+ B cells. *J Clin Invest* **99**, 1614–1627 (1997).
198. Collins, A. M. & Watson, C. T. Immunoglobulin light chain gene rearrangements, receptor editing and the development of a self-tolerant antibody repertoire. *Frontiers in Immunology* **9**, 2249 (2018).

199. vander Heiden, J. A. *et al.* Dysregulation of B Cell Repertoire Formation in Myasthenia Gravis Patients Revealed through Deep Sequencing. *The Journal of Immunology* **198**, 1460–1473 (2017).
200. Panigrahi, A. K. *et al.* RS rearrangement frequency as a marker of receptor editing in lupus and type 1 diabetes. *Journal of Experimental Medicine* **205**, 2985–2994 (2008).
201. Harris, K. E. *et al.* Sequence-Based Discovery Demonstrates That Fixed Light Chain Human Transgenic Rats Produce a Diverse Repertoire of Antigen-Specific Antibodies. *Front Immunol* **9**, (2018).
202. Mcdaniel, J. R., DeKosky, B. J., Tanno, H., Ellington, A. D. & Georgiou, G. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nature Protocols* (2016) doi:10.1038/nprot.2016.024.
203. DeKosky, B. J. *et al.* In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nature Medicine* **21**, 86–91 (2015).
204. Lefranc, M. P. Nomenclature of the Human Immunoglobulin Kappa (IGK) Genes. *Experimental and Clinical Immunogenetics* **18**, 161–174 (2001).
205. Stamatopoulos, K. *et al.* Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood* **106**, 3575–3583 (2005).
206. Jackson, K. J. L. *et al.* Divergent human populations show extensive shared IGK rearrangements in peripheral blood B cells. *Immunogenetics 2011 64:1* **64**, 3–14 (2011).
207. Immune Repertoire Sequencing. https://irweb.irepertoire.com/nir/showsample?sampleid=21743.
208. Immune Repertoire Sequencing. https://irweb.irepertoire.com/nir/showsample?sampleid=21744#.
209. Immune Repertoire Sequencing. https://irweb.irepertoire.com/nir/showsample?sampleid=21745#.
210. Immune Repertoire Sequencing. https://irweb.irepertoire.com/nir/showsample?sampleid=21746#.
211. Liu, H. *et al.* Cross-neutralization of a SARS-CoV-2 antibody to a functionally conserved site is mediated by avidity. *bioRxiv* (2020) doi:10.1101/2020.08.02.233536.
212. Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein. *Science (1979)* **369**, (2020).
213. Brown, J. F. *et al.* Anti-Ebola Virus Antibody Levels in Convalescent Plasma and Viral Load After Plasma Infusion in Patients With Ebola Virus Disease. *J Infect Dis* **218**, 555–562 (2018).
214. Liu, X. *et al.* Neutralizing Aptamers Block S/RBD-ACE2 Interactions and Prevent Host Cell Infection. *Angew Chem Int Ed Engl* **60**, 10273–10278 (2021).
215. Libster, R. *et al.* Early High-Titer Plasma Therapy to Prevent Severe Covid-19 in Older Adults. *New England Journal of Medicine* **384**, 610–618 (2021).

216. Gontu, A. *et al.* Limited window for donation of convalescent plasma with high live-virus neutralizing antibody titers for COVID-19 immunotherapy. *Communications Biology 2021 4:1* **4**, 1–9 (2021).
217. Sui, J. *et al.* Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proc Natl Acad Sci U S A* **101**, 2536–2541 (2004).
218. Grant, T., Rohou, A. & Grigorieff, N. cis TEM, user-friendly software for single-particle image processing. *Elife* **7**, (2018).
219. Xiong, X. *et al.* A thermostable, closed SARS-CoV-2 spike protein trimer. *Nature Structural & Molecular Biology 2020 27:10* **27**, 934–941 (2020).
220. Henderson, R. *et al.* Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nature Structural & Molecular Biology 2020 27:10* **27**, 925–933 (2020).
221. McCallum, M., Walls, A. C., Bowen, J. E., Corti, D. & Veesler, D. Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation. *Nat Struct Mol Biol* **27**, 942–949 (2020).
222. Hsieh, C. L. *et al.* Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501–1505 (2020).
223. Ju, B. *et al.* Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **584**, 115–119 (2020).
224. Cao, Y. *et al.* Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell* **182**, 1–12 (2020).
225. Barnes, C. O. *et al.* SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature 2020 588:7839* **588**, 682–687 (2020).
226. Yuan, M., Liu, H., Wu, N. C. & Wilson, I. A. Recognition of the SARS-CoV-2 receptor binding domain by neutralizing antibodies. *Biochemical and Biophysical Research Communications* **538**, 192–203 (2021).
227. Rogers, T. F. *et al.* Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science (1979)* **369**, 956–963 (2020).
228. Zhou, X. *et al.* Diverse immunoglobulin gene usage and convergent epitope targeting in neutralizing antibody responses to SARS-CoV-2. *Cell Reports* **35**, 109109 (2021).
229. Simpson, R. J. Fragmentation of protein using trypsin. *CSH Protoc* **2006**, pdb.prot4550 (2006).
230. Giron, C. C., Laaksonen, A. & da Silva, F. L. B. On the interactions of the receptor-binding domain of SARS-CoV-1 and SARS-CoV-2 spike proteins with monoclonal antibodies and the receptor ACE2. *Virus Research* **285**, 198021 (2020).
231. Zost, S. J. *et al.* Potently neutralizing human antibodies that block SARS-CoV-2 receptor binding and protect animals. *bioRxiv* 2020.05.22.111005 (2020) doi:10.1101/2020.05.22.111005.

232. Premkumar, L. *et al.* The receptor binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Sci Immunol* **5**, (2020).
233. Vanshylla, K. *et al.* Discovery of ultrapotent broadly neutralizing antibodies from SARS-CoV-2 elite neutralizers. *Cell Host & Microbe* **30**, 69-82.e10 (2022).
234. Schickel, J. N. *et al.* Self-reactive VH4-34–expressing IgG B cells recognize commensal bacteria. *Journal of Experimental Medicine* **214**, 1991–2003 (2017).
235. Bryson, S. *et al.* Structures of Preferred Human IgV Genes–Based Protective Antibodies Identify How Conserved Residues Contact Diverse Antigens and Assign Source of Specificity to CDR3 Loop Variation. *The Journal of Immunology* **196**, 4723–4730 (2016).
236. Thomson, C. A. *et al.* Germline V-genes sculpt the binding site of a family of antibodies neutralizing human cytomegalovirus. *The EMBO Journal* **27**, 2592 (2008).
237. Khurana, S. *et al.* Human antibody repertoire after VSV-Ebola vaccination identifies novel targets and virus-neutralizing IgM antibodies. *Nat Med* **22**, 1439–1447 (2016).
238. Devito, C. *et al.* Human IgM monoclonal antibodies block HIV-transmission to immune cells in cervico-vaginal tissues and across polarized epithelial cells in vitro. *Scientific Reports 2018 8:1* **8**, 1–12 (2018).
239. Klingler, J. *et al.* Role of IgM and IgA Antibodies in the Neutralization of SARS-CoV-2. *medRxiv* (2020) doi:10.1101/2020.08.18.20177303.
240. Melchers, F. *et al.* The surrogate light chain in B-cell development. *Immunol Today* **14**, 60–68 (1993).
241. Xu, L. *et al.* Combinatorial surrobody libraries. *Proc Natl Acad Sci U S A* **105**, 10756–10761 (2008).
242. Sun, L. *et al.* Identification of Mouse and Human Antibody Repertoires by Next-Generation Sequencing. *J Vis Exp* (2019) doi:10.3791/58804.
243. Rettig, T. A., Ward, C., Bye, B. A., Pecaut, M. J. & Chapes, S. K. Characterization of the naive murine antibody repertoire using unamplified high-throughput sequencing. *PLOS ONE* **13**, e0190982 (2018).