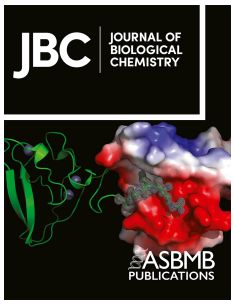


Expanding and Improving Nanobody Repertoires Using a Yeast Display Method: Targeting SARS-CoV-2

F.R. Cross, P. Fridy, N. Ketaren, F. Mast, S. Li, P. Olivier, K. Pecani, B.T. Chait, J.D. Aitchison, M.P. Rout

PII: S0021-9258(23)00086-8

DOI: https://doi.org/10.1016/j.jbc.2023.102954

Reference: JBC 102954

To appear in: *Journal of Biological Chemistry*

Received Date: 3 November 2022

Revised Date: 24 January 2023

Accepted Date: 25 January 2023

Please cite this article as: Cross FR, Fridy P, Ketaren N, Mast F, Li S, Olivier P, Pecani K, Chait BT, Aitchison JD, Rout MP, Expanding and Improving Nanobody Repertoires Using a Yeast Display Method: Targeting SARS-CoV-2, *Journal of Biological Chemistry* (2023), doi: https://doi.org/10.1016/j.jbc.2023.102954.

CRediT Author Statement

F.R. Cross: Conceptualization, Methodology, Software, Investigation, Supervision, Writing, Funding acquisition; P. Fridy, N. Ketaren, F. Mast, S. Li, P. Olivier, K. Pecani: Investigation, Resources; B.T. Chait: Resources, Writing, Funding acquisition; J.D. Aitchison: Supervision, Writing, Funding acquisition; M. P. Rout: Conceptualization, Methodology, Investigation, Supervision, Writing, Funding acquisition.

Expanding and Improving Nanobody Repertoires Using a Yeast Display Method: Targeting SARS-CoV-2

F.R. Cross[1]*, P. Fridy[2], N. Ketaren[2], F. Mast[3], S. Li[3], P. Olivier[3], K. Pecani[1], B.T. Chait[4], J.D. Aitchison[3], M. P. Rout[2]*.

1. Laboratory of Cell Cycle Genetics, The Rockefeller University, New York, New York 10065, 12 USA
2. Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, New York 10065, 12 USA
3. Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, 10 Washington 98109, USA.
4. Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, 14 New York 10065, USA

*Correspondence: rout@rockefeller.edu; fcross@rockefeller.edu

ABSTRACT

COVID-19, caused by the coronavirus SARS-CoV-2, represents a serious worldwide health issue, with continually emerging new variants challenging current therapeutics. One promising alternate therapeutic avenue is represented by nanobodies, small single chain antibodies derived from camelids with numerous advantageous properties and the potential to neutralize the virus. For identification and characterization of a broad spectrum of anti-SARS-CoV-2 Spike nanobodies, we further optimized a yeast display method, leveraging a previously published mass spectrometry based method, using B-cell cDNA from the same immunized animals as a source of V_HH sequences. Yeast display captured many of the sequences identified by the previous approach, as well as many additional sequences that proved to encode a large new repertoire of nanobodies with high affinities and neutralization activities against different SARS-CoV-2 variants. We evaluated DNA shuffling applied to the three complementarity-determining regions (CDRs) of antiviral nanobodies. The results suggested a surprising degree of modularity to CDR function. Importantly, the yeast display approach applied to nanobody libraries from immunized animals allows parallel interrogation of a vast number of nanobodies. For example, we employed a modified yeast display to carry out massively parallel epitope binning. The current yeast display approach proved comparable in efficiency and specificity to the MS-based approach, while requiring none of the infrastructure and expertise required for that approach, making these highly complementary approaches that together appear to comprehensively explore the paratope space. The larger repertoires produced maximize the likelihood of discovering broadly specific reagents and those that powerfully synergize in mixtures.

INTRODUCTION

The COVID-19 pandemic, caused by the sarbecovirus SARS-CoV-2, has had a profound global impact the likes of which has not been seen in more than a century. The remarkably rapid development and distribution of vaccines undoubtedly saved many millions of lives (1,2); nevertheless, at the time of writing, mortality estimates range from 10 – 20 million, with additional profound long-lasting health impacts for many survivors (3,4). The disease appears to be

transitioning to an endemic phase, thus presenting a serious worldwide health problem for the foreseeable future (5,6) and demanding a large-scale ongoing implementation of new prophylactics and therapeutics.

Major therapeutic strategies have utilized antibodies directed against the major Spike (S) surface envelope glycoprotein of the SARS-CoV-2 virion, various fragments of which are also the immunogens for most vaccines. Spike is a homotrimer of an extensively glycosylated ~200 kDa protein composed of two major domains: S1, which contains the host receptor binding domain (RBD) that targets the angiotensin-converting enzyme 2 (ACE2) surface receptor on host cells; and S2, which upon host cell binding undergoes major conformational changes to enable viral – host membrane fusion, resulting in virus entry into the cytoplasm (7-12). Thus, antibodies that target Spike, and particularly RBD – generated through vaccination or exogenously introduced - have the potential to block viral binding and entry into the cells of the host. Unfortunately, the continuing emergence of new SARS-CoV-2 Variants of Concern (VoCs; Alpha, Beta, Gamma, Delta, Omicron and subvariants) presents a significant barrier to attaining complete control of COVID-19. These VoCs usually have many Spike mutations (especially the RBD), and thus are relatively poorly neutralized by current vaccines and antibody therapies (13,14). For instance, monoclonal antibodies (mAbs) have proven to be an effective therapeutic strategy, though sensitive to emerging variants (15,16). Moreover, mAbs are limited by challenges in the ease and cost of their large-scale manufacturing, distribution, and intravenous administration (17).

A promising alternative to mAbs is a particular class of single domain antibodies termed “nanobodies”. Nanobodies are “mini-antibodies”, some $1/10^{th}$ the size of regular IgGs, derived from the variable domain (V_HH) of variant heavy chain-only IgGs (HCAbs) found in camelids (e.g. llamas). Each nanobody molecule is constructed of a single Ig fold, consisting of four framework regions (FRs) that intersperse and orient three complementarity determining regions (CDRs) that form the nanobody paratope (18). These regions are similar to FRs and CDRs of conventional antibodies (19). As with conventional antibodies, CDR3 is formed by VDJ recombination of germline DNA; CDR1 and CDR2 come from the germline V region, and all three CDRs are then subject to somatic hypermutation, with selection for improved binding affinity to antigens (19).

Nanobodies have several attractive advantages over mAbs, including: extremely fast on-rates leading to high overall affinities (20-22); characteristics of small molecules in terms of higher tissue penetration and accessibility to regions of Spike not accessible to the larger mAbs or occluded by glycosylation (20,21,23,24), greatly enhancing their potential to synergize in combination, a profound advantage they have over often poorly-synergizing conventional antibodies (20,25,26). They can be readily engineered, including humanization to minimize immunogenicity; they are highly denaturation-resistant, giving them long shelf lives and making them suitable for a broader range of delivery methods (e.g., via nebulization directly into lungs) (27,28); and very low cost of production in bacterial or yeast expression systems (20,29-32). Like mAbs, nanobody binding can be disrupted by mutations in VoCs. However, by producing large repertoires targeting the entirety of Spike, we have made many nanobodies that are resistant to the major VoCs (20). Moreover, their small size allows a high multiplicity of binding to Spike, which, as we have shown, allows the judicious selection of nanobody cocktails that can pack around different epitopes as well as all three copies of the same epitope; these generate a tremendous synergistic effect that can also be highly VoC resistant, as the virus must now evolve multiple simultaneous mutations to evade such cocktails (20).

To produce such valuable nanobody repertoires, we have previously employed a mass spectrometry-based approach (20,22) in which we immunized llamas with Spike constructs, taking advantage of powerful natural affinity maturation processes *in vivo* (33), and then

performed high-throughput DNA sequencing of V_HH libraries PCR-amplified from marrow lymphocyte cDNA from the immunized llamas in combination with mass spectrometric (MS) identification of high-affinity V_HH regions derived from the serum of the same animal. Computational matching of MS-sequenced peptides to V_HH cDNA sequences allowed high-confidence identification of sequences encoding high-affinity nanobodies. Genes encoding nanobodies were synthesized and expressed in bacteria, and nanobodies were purified, and characterized for their specificity and affinity. However, these V_HH cDNA libraries also represent a resource that can be tapped for an orthogonal approach for nanobody production, employing display screening methods instead (34-37). Potentially, this approach could discover additional nanobodies to enrich our repertoires, and also serve as a convenient platform to explore the specificity and VoC sensitivity of a large number of nanobodies in parallel. Recently, a robust and efficient yeast display method was designed and validated specifically for screening nanobodies, in particular against Spike (34,38). In that work, a synthetic nanobody library was employed. Here, we chose to evaluate instead the cDNA library made from immunized llamas and already tested and mined by the MS-based approach (20), which we transferred into the nanobody display vector (34). We tuned the display approach and selected large repertoires of nanobodies that were specific for different domains of Spike and contained members that displayed high affinity and resistance to VoCs, and also employed DNA shuffling of the CDRs to generate variants with novel VoC specificities. We show that the yeast display method, either on its own or in parallel with the mass spectrometric method, can generate large nanobody repertoires with high potential as therapeutics, in this case against COVID-19.

RESULTS & DISCUSSION

Optimization of Yeast Display Screening

We show our general pipeline design in Fig. 1. We used a well-characterized nanobody display vector (34), which gave efficient and selective binding of specific nanobodies to diverse targets in our hands (GFP, and the S1, RBD and S2 regions of Spike). Instead of published approaches such as rounds of fluorescence-activated cell sorting (FACS) (39), or Miltenyi magnetic bead and biotin-based purification followed by FACS (34), we employed Dynabeads coupled to antigen for all selection steps (Fig. 1, inset). For most of these initial experiments we generated a set of four anti-GFP nanobodies in the display vector (22) with a broad range of known affinities, testing for binding of yeast expressing these to GFP-conjugated Dynabeads, as well as an irrelevant non-specific nanobody control. We also cloned three previously characterized anti-Spike nanobodies (20) into the yeast display vector to test binding to various Spike domain-conjugated Dynabeads. We used all these well-characterized nanobodies as benchmarks to optimize conditions in terms of simplicity, specificity, and yield, trying for an optimal combination of high binding of specific nanobody-expressing yeast and low binding of control yeast carrying just the vector. We tested a range of binding buffers and conditions to optimize specific binding to these beads, taking advantage of the extreme robustness of yeast to even harsh binding conditions including high salt, as well as acidic and detergent washes. We found that a straightforward buffer, related to ones used for immunofluorescence microscopy and immunoblotting, gave excellent signal-to-background with these benchmarks (see Methods). Often, for specific binding, multiple beads were attached to each yeast cell, whereas no beads would be seen bound in the nonspecific control (Figs. 1, 2A). Magnetic isolation of the Dynabead-binding yeast was found to give the highest yields while maintaining efficient removal of nonspecific yeast when the magnet was placed at a distance from the yeast washing suspension such that it took several minutes to fully harvest the beads; closer placement of the magnet led to loss of cells during the repeated washes. Care was taken on each wash step so that the beads were fully resuspended with minimal displacement of specifically-attached yeast from them. We found that 4-5 wash steps were sufficient to achieve a 100-1000-fold enrichment of specific binders over the control in a single round of isolation, which was sufficient for clear identification of specific binders by comparison of sequence counts before and after binding (Fig. S1).

Determination of the Roles of Nanobody Affinity and Competition During Screening

Competition has the potential to dramatically limit the repertoire, so we sought conditions where we could isolate high affinity nanobodies, but not at the expense of competition among yeast - an effect that could be exacerbated by avidity effects. Therefore, we used our model nanobody display strains to determine the avidity effects of the many nanobodies displayed per yeast cell, and competition between yeast cells displaying nanobodies of differing affinities for the same antigen. In order to distinguish the cells of strains displaying nanobodies of differing affinities, we covalently stained cell walls with vital fluorescent dyes (40) (Fig. 2A). To examine the effects of competition, we mixed equal cell numbers of two strains of yeast each uniquely dyed and expressing one of two different anti-GFP nanobodies of differing affinities (22), and also unlabeled control cells expressing an irrelevant nanobody. These three cell populations were in ratios of 50:50:0 or 1:1:98 respectively, all at the same final cell density and volume; the amount of GFP-Dynabeads was also held constant. Initial experiments suggested that in the 50:50:0 condition, beads might be limiting, since we observed depletion of free beads not bound to yeast in the final harvest (see Fig. 1 for an example of bead depletion). In contrast, at the 1:1:98 ratios there appeared to be abundant beads remaining after binding and washes. Therefore, we call the experiment carried out at 50:50:0 ‘competitive’ and the 1:1:98 ‘noncompetitive’.

To assess the effect of competition as a function of binding affinity, we quantified the variability of yield across all possible pairwise binding experiments (Fig. 2B). The yield of yeast displaying low-affinity nanobodies was highly variable, because when competing with yeast displaying successively higher affinity nanobodies the yield of the former was increasingly depleted. This effect was far stronger in the competitive versus noncompetitive conditions. In contrast, yeast expressing high-affinity nanobodies displayed little variation in yield even under competitive conditions, because they 'won' the competition. These results clearly demonstrate that the yeast display system is sensitive to the monomer binding affinity of the displayed nanobodies. We find this result surprising due to avidity considerations, but the effect seems quite clear, and was further validated in testing anti-Spike nanobodies of known affinities (see below). This effect would become exacerbated under the increasingly competitive conditions that will arise through repeated rounds of panning characteristic of display methods, leading to isolation of the most competitive clones at the expense of other less competitive but still potentially valuable high affinity clones. Thus, we instead sought to take advantage of (i) starting with a hyper-immune animal and (ii) the high signal-to-background of our optimized panning conditions to limit our screens to just two rounds of panning, which we found to be sufficient under these conditions to identify strong positive clones that outcompete non-specific or low affinity clones, not at the expense of outcompeting other valid high affinity clones.

Examination of the Sequence Diversity of the Yeast Display Libraries.

We used gap-repair in yeast to clone into the yeast display vector B-cell V_HH cDNA from two llamas immunized with Spike proteins, 5094 ('Marley') and 7704 ('Rocky') (20). We obtained libraries of respectively ~5 x 10^6 and ~1.5 x 10^7 independent clones. We carried out next-generation sequencing on the inserts amplified from the transformed yeast. We observed highly diverse sequences, especially in the known hypervariable regions CDR1,2,3, and also considerable though lesser variation in the framework regions (Fig. 3A, B). We noted that the library contained many 'families' of closely related but distinct sequences. Variability within families was reduced compared to variability in the library overall, but was still substantial even in the framework regions (Fig. 3C). Positions of variation in framework regions within families were similar to the variable positions in the overall library. We believe this is likely due to these variable positions occupying loop or surface positions far from the antigen binding site in the nanobody structure, since this has been noted previously for positions of nanobody variation (41,42). This, in turn, suggests that diversity within these sequence families was generated under selection in the llama, implying that this variation is largely due to somatic hypermutation after 'founding' of the family by V-D-J recombination (see Introduction) rather than artifactually introduced during cloning and sequencing. The multiple related sequences in these families effectively provide a large number of biological replicates for binding experiments, as shown below.

Validation of the Yeast Display Method with Biochemically Identified Nanobodies

We screened the yeast display libraries from the two Spike-immunized llamas using the protocol optimized with the control strains (above; see Methods). Dynabeads coupled to S1, RBD or S2 each bound to approximately 1% of the clones in the yeast library. To evaluate this binding reaction, we amplified and sequenced the selected clones, and compared the recovery of sequences to their representation in the unselected library shown in Fig.3. To simplify the analysis as well as to focus on the major paratope (i.e., epitope-binding) region of each nanobody, we reduced each sequence to a representation consisting solely of its three CDR regions (CDRs 1,2,3) (we refer to this as the 'CDR string'). We plotted the log2 of read-count for each CDR string

in the unselected library vs. the log2 of read count in the selected libraries (in all cases after correction to reads per million total sequenced). We selected over two rounds with the viral antigens coupled to Dynabeads ('1X' and '2X'). Fig. S1 shows the behavior of the entire library. We observe two broad lines of plotted points, one showing strong enrichment (ratio of selected to unselected much greater than 1, shown by the blue line) and one showing strong depletion (below the blue line). Both broad lines have slopes of ~1, indicating a first classification of sequences into specific and non-specific binders, as indicated in red in Fig. S1. The rising slope in both cases is due simply to more recovery of sequences that are more abundant in the initial library (showing why it is essential to have sequences for the initial as well as the selected libraries). Similar plots were obtained for all three viral antigens (Fig. S1) and both libraries (Fig. S2).

Previously, we identified 374 unique CDR3s by using a biochemical and mass spectrometry-based method and from them cloned and expressed 116 high-affinity anti-Spike nanobodies (20), here termed 'MS positives' since the yeast display libraries here employed the same nanobody cDNA used in that study. We therefore used the biochemically characterized nanobodies as fiduciary markers to analyze their behavior when displayed by yeast. Importantly, these nanobodies bound with the expected specificity when expressed in yeast. Similar results were obtained with both animals (Figs. 4A and S2). For 5094, to increase representation due to lower overall anti-Spike nanobody levels in this animal as described previously (20), we also plotted close relatives (no more than 20% sequence divergence in any CDR) of the MS positives. The screen clearly distinguished specifically binding and non-binding clones; for example, biochemically defined S2-specific nanobodies bound to S2 beads but not S1 or RBD beads when expressed in yeast, and vice versa. We plotted enrichment of these MS positive sequences after both 1 and 2 rounds of selection against their measured K_Ds (20). While the relationship was noisy, a statistically significant negative slope was observed (Fig. 4B), especially for two rounds of selection (likely to be more competitive conditions based on our analysis of anti-GFPs above), once again confirming the competitive and affinity-sensitive nature of the screen. Collectively, these results validate the screen's ability to identify large numbers of *bona fide* high affinity Spike-binding nanobodies, as well as revealing two other key behaviors. First, members within a related family behave similarly, in terms of relative enrichment and specificity during panning (e.g., Figs. S2 and S3; see also following figures). Second, some families came to dominate the final panned populations, particularly after two rounds of selection; for example, clones with a CDR3 containing the sequence string "GANAAH" made up more than 90% of the recovered sequences from 5094 with RBD- or S1-Dynabeads (Fig. S2). However, overall diversity of the positive families was still preserved despite distortion of representation.

We were also interested to employ this same benchmark to evaluate previously described panning methods (34,38), in which after a first round of magnetic bead purification, subsequent panning rounds utilize flow cytometry after binding of fluorescent RBD (Fig. S3). We found the first step of the previously described method, with magnetic bead-based enrichment of binders, gave similar results to our procedure. In our hands, the second flow-cytometry based step very effectively eliminated non-specific binders; however, we also noted that the broad representation observed with our method was lost, and the recovered clones were strongly dominated by a single sequence (with identical CDRs to the MS positive S1-RBD-38 (20)). The stringent removal of non-specific binders is likely critical for screening naive libraries as in (34,38), but we find that it is not required in our context starting with V_HH cDNA from hyperimmunized llamas, where our method clearly preserves the high diversity present in the original cDNA.

Screening for New Families of Anti-Spike Nanobodies

We next analyzed the screens described above and in Fig. 4 to discover novel nanobodies not previously identified by the mass spectrometric method, focusing mainly on the more diverse repertoire of animal 7704 (20,43). We selected for analysis, using a polygon function (see Methods), the ~200 most abundant and highly enriched clones from the first round of antigen selection specifically for that given antigen, and the CDR families for those ~200 clones were also selected (dark and light colors respectively; Fig. 5A, upper row). These were found to comprise the most highly enriched clones from the second round of antigen selection, and preserved their antigen specificities (Fig. 5A, lower row).

We examined the sequence diversity and content of these positive classes by examination of neighbor-joining trees (Fig. 5B). While many of the sequences were similar or identical to the previously described nanobodies from this llama identified by the MS-based approach (20), we also observed many new sequences. We used CDR3 as our benchmark for defining a given family, because its generation by VDJ recombination is the unique clonal event founding a large cluster of related sequences by somatic hypermutation. The estimates for the number of families that recognize each spike domain vary depending on the sequence parameters used to define each family, the antigen in question, the cutoff for enrichment used during a panning round, the cutoff for minimum readcount (to exclude spurious sequencing / PCR errors), and the minimum size of a family. We used similarities within CDR3 to define families, with each family being defined as being made of members related to each other within a certain value or less of sequence identity. For example, for animal 7704 with the more inclusive parameters of 70% or more CDR3 sequence identity and a minimum enrichment of twofold, we found 259 S2-specific families, 90 RBD-specific families also recognizing S1, and 60 S1 (non-RBD) families. Interestingly, there are a further 106 families, mostly low readcount, that recognize RBD but not S1, presumably antigens buried in the full Spike protein. By any count, however, it is clear that a very large number of anti-Spike families - often over one hundred per domain - can be identified by this approach.

We constructed a neighbor joining tree, and though of course not a ‘phylogeny’ tree in the strictest sense, the nodes of the tree nevertheless correspond to the sequence families discussed above (see also (20)). Major representative families are denoted in Fig. 5B with a Roman numeral and Seqlogos (44) were constructed for these families (Fig. 5C). It can be seen that the CDR1,2,3 sequences are very diverse between families, but largely contain only minor variations within families. Moreover, a given family shows absolute antigen specificity, such that e.g. an RBD-specific family is not found enriching in an S2 screen. The S1-non-RBD clones are dominated by members of one family, defined by its ‘IAQY’ consensus CDR3 sequence (Fig. 5B,C; family (i)). This family was missed in the MS-based approach (20), possibly because the CDR3 was too small for reliable peptide identification. However, numerous other new S1-specific families are present, for example the “RGLGRGLGFY” CDR3 consensus sequence (Fig 5B, family (ii)). By contrast, and in agreement with its antigenic nature (12,20,45,46), the RBD domain isolated a large and diverse set of families. One of the largest families (Fig 5B, family (v)) contains the consensus CDR3 “TVDAQSDY”, which is also found in the mass spectrometrically identified nanobody S1-RBD-38. Two large families identified here contain divergent relatives among the MS-identified nanobodies (20), the ‘’LRSRFNAAAWTTEAAFDY’’ (previous MS-identified S1-RBD-6 and S1-RBD-31) and ‘YERLAWDTSTY’’ families (previous MS-identified S1-RBD-35), the remaining four indicated families being completely novel. The S2-specific clones were also very diverse and not dominated by any single family (Fig. 5B,C), with limited overlap with the mass spectrometrically identified clones (20). Collectively, the screening method identified a large number of new clones recognizing different Spike domains. It then became important to determine if these new clones had high affinity binding and strong antiviral activity when expressed as monomers.

Testing the Library against the Major VoCs Delta and Omicron

One of the greatest challenges to managing COVID-19 is the ability of the SARS-CoV-2 virus to mutate into new VoCs that can resist prevalent vaccines and therapeutics. An advantage of generating large repertoires of nanobodies is that one maximizes the likelihood of finding VoC-resistant, broadly specific nanobodies (20,47-49). At the time of writing, two variants, and derivatives thereof, remain of significant concern, Delta and Omicron. We wanted to evaluate the utility of yeast display for analyzing the sensitivity of our nanobody repertoire to these variants, focusing on the RBD domain as it is the region of largest mutational variation (50,51), and the 7704 animal as it has the largest representation. We conjugated recombinant Delta and Omicron RBD to Dynabeads and tested for binding to the yeast library compared to the original SARS-CoV-2 strain's RBD (two rounds of selection, with sequence analysis after each). Remarkably, overall, we observed that most sequences bound well to both variants (Fig. 6A; Fig. 7). For example, the large 'TVDAQSDY' (Fig. 5C, (v)) family binds comparably to the original SARS-CoV-2 and both variants. The ''LRSRFNAAAWTTEAAFDY' ('NAAAW') (Fig. 5C, (iv)) family binds comparably to the original SARS-CoV-2 and the Delta VoC RBDs, and appears collectively slightly weaker against the Omicron RBD. The 'YERLAWDTSTY' ('YERLAWD') (Fig. 5C, (iii)) family binds well to the original SARS-CoV-2 but binding is essentially eliminated to both variants. The 'IIDDYGVQY' ('IIDDY') (Fig. 5C, (vi)) family (Fig. 6A, 'IIDDY') and a family not indicated on Fig. 5 but comprising a family characterized by a 'TADLYSDY' ('TADLY') CDR3 sequence binds well to original SARS-CoV-2 and the Omicron variant but more weakly to the Delta variant, especially after two rounds of selection. These families contain considerable sequence diversity within them (Fig. 5). There are many other unrelated clones (often with lower representation in the library) that exhibit similar behaviors (Fig. 6A), greatly expanding the useful repertoire. A similar behavior was seen for the 'MS positives' in our screening (Fig. S4). This screening method is therefore potentially a rapid and straightforward way to further characterize the library for the VoC-specific sensitivities of the positive clones.

Testing the Modularity of CDRs by DNA Shuffling

DNA shuffling is an established *in vitro* method for improvement of binding or catalytic activity, and has previously been applied to nanobodies with recombination between CDRs, resulting in significant improvements of binding noted in the progeny (52-54). Typically it is applied to a library of randomly point-mutagenized sequences that have been selected for improved activity, from which starting point splice-overlap-extension (SOE) PCR is carried out to produce mix-and-match recombinants. Two advantages ensue from this method compared to simple clonal descent by cycles of mutagenesis and selection. First, deleterious mutations hitchhiking with selected mutations are readily crossed away; second, combinations of positive mutations in different regions can be combined in a single jump through sequence space that might be highly unlikely to occur by single-mutation steps. Improved activity, and even new biological activities entirely lacking in the starting material, can potentially be found in the products. It is interesting that while point mutagenesis is a biological strategy naturally used in the process of somatic hypermutation to improve antibody affinity (see Introduction), DNA shuffling is much rarer in natural biological systems. VDJ recombination shares some features, but with the critical difference that only a single round of shuffling occurs rather than multiple rounds interleaved with the recombinations. Neither the extremely high density of recombination joins that can be attained by *in vitro* shuffling, and its highly multiparental nature, are shared by natural biological systems, to our knowledge.

We started with the library selected on SARS-CoV-2 RBD, and carried out SOE recombining CDRs 1, 2 and 3 at random from that library (Fig. S5). The advantage of this approach is that the

specificity of the large number of input positive sequences is known, unlike in previous approaches (52-54). We bound the recombinant library (containing about 10^6 members) to RBD-Dynabeads, and carried out next-generation sequencing on both the input shuffled library and the selected sequences. This allowed us to quantitatively evaluate the degree of enrichment over the shuffled unselected library.

We expected that incompatibility of CDRs from entirely unrelated nanobodies would result in a large majority of the shuffled progeny having highly reduced binding activity; however, this was not observed. A high proportion of the shuffled library could bind to RBD-Dynabeads at least to some extent. We organized the data by first examining the fate of specific CDR3 sequences, examining what CDR1 and CDR2 sequences were associated with high or low binding to RBD (either the original sequence (henceforth here termed Original), or the Delta and Omicron variants, as described already for the initial library (Fig. 6A)). In general, different CDR3 groups retained similar overall specificity of binding to Delta and Omicron to what was seen in the corresponding native families (Fig. 6). However, a notable exception was observed with the 'YERLAWD' CDR3. The native YERLAWD family bound efficiently to Original but not to Delta or Omicron RBD. The shuffled YERLAWD group, in contrast, contained abundant members that bound equally well to Original and to Delta RBD, while remaining almost completely defective in Omicron binding. Examination of the sequences associated with this high Delta binding revealed specific enrichment of sequences highly similar to the native 'NAAAW' CDR1 and CDR2, recombined with the 'YERLAWD' CDR3 (Fig. 6B). One explanation of this observation could be that the native 'NAAAW' family CDR1 and CDR2 have specific ability to bind to Delta RBD, independent of CDR3 content. However, direct examination of all shuffled products containing these CDR1,2 sequences shows that many fail to bind Delta RBD; binding requires specific CDR3s as well. So, fusion of CDR1,2 from the native NAAAW family to the YERLAWD CDR3 may create a fusion with effective multipoint attachment to the Delta RBD.

We biochemically characterized this apparent 'rescue' of 'YERLAWD' binding to Delta RBD by expressing four distinct 'NAAAW' native parents and two 'YERLAWD' native parents, and recombinants between them, as monomeric recombinant nanobodies (Fig. 7B; see also next section). As expected from the yeast display data (Fig. 6), the two 'YERLAWD' parent nanobodies failed to bind Delta RBD while binding strongly to the Original RBD; and, the four 'NAAAW' parent nanobodies bound strongly to RBD from both strains (Fig. 7B). However, these hybrid nanobodies bound less well to Original and Delta RBD than either of their parents (with approximately equal affinities to Original and Delta), failed to neutralize (not shown), and showed denaturation at relatively low temperatures (Fig. 7C). These defects could simply reflect minor folding incompatibilities in the shuffled construct. Nevertheless, the shuffling results overall suggest significant modularity in CDR function, and could provide a novel avenue to new specificities which could be useful in diverse contexts.

Neutralization and Biophysical Characterization of the Isolated Nanobodies on VoCs

We selected 30 abundant sequences that represented a cross-section of the major families (Fig. 5; Supplementary Table 1), most of which were distinct from clones previously isolated (20). We used the bacterially expressed and purified monomers to determine binding affinities and neutralization capability, using assays we have employed previously (20); we also tested binding affinities and neutralization against Delta and Omicron variants (Fig. 7).

With regards to neutralization, of the 21 S1-targeting nanobodies tested (both RBD and non-RBD binding), all 21 were capable of neutralizing Original, while 10 were still able to neutralize both Delta and Omicron, although with reduced activity against the variants in some cases. We also

tested 9 anti-S2 nanobodies for neutralization of Original; some neutralization activity was observed, although less efficient than for the better of the anti-S1 nanobodies. Such reduced neutralization was also true of our previously characterized anti-S2 nanobodies (20). A subset of the anti-S1 nanobodies were tested for binding affinity against recombinant S1 or RBD from either Original, Delta or Omicron (S1 non-RBD nanobodies were not tested against omicron) (Fig. 7B); all bound strongly to Original, displaying affinities in the nM - pM range. Two of these failed to bind only Delta, two failed to bind only Omicron and CoV2-YD-33 and CoV2-YD-34 failed to bind both variants. Aside from CoV2-YD-10, all were in agreement with their failure to neutralize Delta, omicron or both strains, and some had a moderate to strong reduced binding to Delta and Omicron, again correlating approximately with their reduced neutralization for that strain (Fig. 7A). CoV2-YD-10 neutralizes yet shows no binding to the RBD of omicron using SPR. It is possible the binding site of the nanobody may be slightly truncated in the RBD construct used for SPR resulting in the no binding result. Lastly, all showed a moderate to strong degree of thermal stability, typical of nanobodies (Fig. 7C) (21).

Overall, the nanobody binding in the yeast display screening (Fig. 6) correlates reasonably well with that seen in the biochemical assay of the corresponding expressed nanobody (Fig. 7B). Thus, the differential affinities of the 'YERLAWD' and 'NAAAW' nanobodies for Delta RBD in both yeast display screening (Fig. 6) and in the biochemical assay of the corresponding expressed nanobody (Fig. 7B) agree, as discussed in the previous section. 'LAYVT' (CoV2-YD-7) shows no significant loss of affinity for either Delta or Omicron RBD in both yeast display screening and in the biochemical assay of the corresponding expressed nanobody. 'TALLS' (CoV2-YD-6) show partial loss of affinity for Delta RBD while retaining Omicron affinity in both yeast display screening and in the biochemical assay of the corresponding expressed nanobody. 'TADLY' (CoV-YD-9) shows complete loss of affinity for Delta RBD in both yeast display screening and in the biochemical assay of the corresponding expressed nanobody, while again retaining Omicron affinity in both assays. One nanobody seems something of an exception: 'IIDDY' (CoV2-YD-8) has lost binding for Omicron RBD in the biochemical assay but no obvious reduction in the yeast display screening; perhaps the avidity effect of the display method compensates for the loss of affinity of the monomeric nanobody (20). For the bulk of the new nanobodies tested, the magnitudes of their binding affinities, neutralization potential, and thermal stabilities were comparable to the nanobodies characterized in (20). Collectively, these similar patterns of behaviors of the selected representatives of this new yeast display repertoire to that of the previously published MS-identified repertoire (20) strongly suggests that the same modifications found in that work to powerfully enhance affinity or neutralization activity - multimerization and synergistic mixtures - will be equally applicable to this current repertoire.

Epitope Mapping of the Nanobody Repertoire

Another important characterization of any nanobody repertoire is to determine the different epitopes being recognized by each nanobody, as in the case of anti-SARS-CoV-2 Spike nanobodies, exploration of a larger epitope space increases the likelihood of discovering variant resistant, strongly neutralizing nanobodies (20). This is usually done by 'epitope binning': finding classes of nanobodies that reciprocally inhibit each others' binding due to competition for the same epitope. Epitope binning is generally carried out by one-on-one competitions between pairs of nanobodies; therefore, the number of assays scales with the square of the number of nanobodies to test. It occurred to us that we could use the yeast display assay to determine nanobody sequences in a given epitope bin across the entire library in a single binding experiment. We tested this idea in a preliminary experiment in which we saturated RBD-Dynabeads with three different nanobodies, each previously shown to bind to a distinct epitope (20). We then tested those beads for binding to yeast cultures, each expressing a single nanobody

in a known biochemically defined epitope bin. Prior nanobody binding quantitatively blocked binding of yeast expressing nanobodies in the same epitope bin, but had essentially no effect on binding of yeast expressing nanobodies known to be in different epitope bins (Fig. S6), thus validating the approach.

To carry out parallel epitope binning for the entire yeast display library, we followed conditions from this preliminary experiment: RBD-Dynabeads were saturated with 7 different nanobodies, each determined to represent 7 epitope classes using an integrative mapping approach which incorporated epitope binning by competitive binding, then subdivided the biochemical epitope bins based on escape mutant and crosslinking-mass spectrometry data (20) (we will refer to these subdivided bins as 'epitope classes'). As with biochemically defined epitope bins, the classes may overlap on the RBD surface. We also attempted to block with the soluble extracellular domain from ACE2. The blocked beads were used to select binders from a library of 2-times-selected RBD binders (Figs. 4, 5). V_HH sequences from the bound population were determined, and the read counts of the sequences bound to RBD beads blocked with each nanobody were determined. The read count recovered from the blocked beads was divided by the read count from the unblocked beads, and the resulting ratios hierarchically clustered (Fig. 8). These 7 epitope classes were selected to collectively encompass essentially all of the available RBD surface (20). Consistent with this, the majority of nanobodies in our population are inhibited by blocking the RBD beads with at least one of the seven nanobody classes (Fig. 8). A minority of nanobodies were not so inhibited and may represent new epitope class(es).

Many nanobodies fall into more than one epitope class, as defined by this assay. Thus, most nanobodies in class #1 are also in class #2, and vice versa; and a similar mutuality is seen between classes #3 and #4, which in turn contains a smaller subgroup that is also found in class #5 (Fig. 8). The positions of the 'founding' epitopes for these classes was estimated previously from MS cross-linking data and escape mutants (20). Examination of the estimated position of these epitopes on RBD indeed indicates that there is significant overlap or adjacency between #1 and #2, and between #3, #4 (and even #5 or #6), consistent with steric clashes that could lead to the class overlaps observed (Fig. 8) (20).

We checked the fidelity of this method by examining behaviors of nanobodies previously identified and classified (20). In total, that work placed 17 nanobodies in one of these seven epitope classes (20). Of those 17, 11 were found in the library; of those 11, 9 exhibited the expected result of specific prevention of binding to RBD beads blocked with the nanobody defining the epitope class. The remaining 2 were less clear, with a related sequence found to one (S1-RBD-21) showing some inhibition in epitope class #6 though previously assigned to #4, and the other (S1-RBD-23) showing no inhibition though previously assigned to class #5.

We further dissected nanobody binding behaviors in this assay by once again following particular nanobody families as defined by their CDRs (as above; Figs. 5, 6). The sequence variants within families generally fell together on the clustergram, which was generated sequence-blind, based solely on the binding behavior in the 7 blocked populations. This result supports the similar behavior of almost all the sequence variants assigned to the families. For example, essentially all of the sequences in the 'YERLAWD' CDR3 family were specifically blocked by #6, and essentially all of the sequences in the 'NAAAW' and 'LAYVT' CDR3 families (Figs. 5, 6) were specifically blocked by #7 (Fig. 8). Many other smaller CDR3 families also displayed similar behaviors. Interestingly, classes #1 to #4 did not singly block any significant CDR3 families, but rather acted to block them in different combinations. Thus, many of the CDR3 families were blocked by both #3 and #4, of note being the large CDR3 family characterized by the starting sequence TVDAQ; however, there appears to be a range of behaviors in this large class (Fig. 8; Fig. S7). So, some

CDR3 families are essentially exclusively blocked by #3 and #4, such as those CDR3 sequences characterized by CDR3s starting with ARDD, ARNQ, and WRYF (Fig. S7, group C). A number of families are similarly strongly blocked by #3 and #4, but also partially blocked (to lesser or greater extents) by #5 (Fig. 8; Fig. S7, group A); these include the large TVDAQ, TALLS, and AAHVN CDR3 starting sequence families (Fig. 8; see also Figs. 5, 6). Then there are those nanobody CDR3 families almost equally strongly blocked by #3, #4 and #5, including S1-RBD-43 and related nanobodies (characterized by the CDR3 starting sequence AGHV) (Fig. S7, group B). Lastly, there are nanobodies that are blocked relatively equally by classes #3, #4, and #6, notably those with CDR3 sequences starting with VDLAP or ASKTT (Fig. 8, Fig. S7; group D). Thus, the epitope classifications are not absolute: there are varying degrees of relative inhibition for each nanobody compared with its neighbors on the plot that indicate there are a very large number of discrete though frequently overlapping epitopes recognized by the population. In contrast, epitope classes that were deduced to be far apart on the RBD surface, such as #6 and #7, exhibit little or no overlap of depletion (intriguingly, analysis of crossover recombinants breaks this rule (Figure S8) for unknown reasons). Within these data are results that also serve to highlight new families not discussed above, some with similar epitope class behaviors as the more common families, and also many that show new behaviors - and among them are also MS-identified with previously unknown specificity, such as close relatives of S1-RBD-16, which is blocked by both #5 and #7 epitope classes. These behaviors suggest that the population of nanobodies effectively ‘scans’ the available epitope space.

CONCLUSIONS

We show that the version of a yeast display method presented here is capable of generating a large repertoire of high affinity nanobodies. A marked advantage of the current method is its robustness, simplicity and relatively low cost in time and resources; it avoids the need for complex and expensive instrumentation and associated necessary expertise. We contend that any laboratory with standard resources and skill sets can readily adapt this method to generate nanobodies against various targets; commercial camelid and sequencing resources can easily be remotely accessed, as they were for this work.

We also show three further adaptations of this method. The first is its use in combination with our previous biochemical / mass spectrometric methods (20,22). We show that in this way, one can significantly increase the total nanobody repertoire. Alternatively, we propose that one can adopt a “two factor” identification approach - only selecting nanobodies that are positive in both the biochemical / mass spectrometric and yeast display methods, these being virtually certain to be true positives with high affinities. Large repertoires are extremely advantageous to maximize epitope space, affinity, the likelihood of obtaining a desired biological activity such as viral neutralization, and candidates for advantageous synergistic mixture or oligomers of nanobodies. The second is the ability to shuffle between different versions of each CDR, which may in certain cases generate nanobodies with new binding behaviors and allows exploration of CDR modularity across the library. The third is that the method can be adapted to allow massively parallel epitope binning. Epitope binning is conventionally and painstakingly performed one nanobody pair at a time. Our method for parallelized epitope binning, in contrast, allows testing many thousands of candidates at once against each known nanobody epitope. This is a necessary step on the road to generating fully characterized nanobodies, including those with diagnostic or therapeutic potential. Indeed here, we have generated many new nanobodies with strong neutralization activity that may be further adapted into treatments for the continuing fight against COVID-19 (20).

ACKNOWLEDGEMENTS

We thank Aashish Manglik and his coworkers for their kind gift of the yeast display vector. We would like to thank the Rockefeller University Genomics Resource Center Core Facility (RRID:SCR_020986) for expert generation of sequence data. We thank members of the Cross, Rout, Chait and Aitchison laboratories for their help. This work was supported by the Robertson Therapeutic Development Fund and the NIH (P41 GM109824).

EXPERIMENTAL PROCEDURES

Library construction

Starting with B cell cDNA from the same immunized llamas described previously (20), we amplified V_HH sequences with oligos providing flanking homology for cloning into the yeast display vector (34). We carried out gap-repair using a high-efficiency yeast transformation method (55),which in our hands yielded a maximum efficiency of colony recovery of ~1.5 * 10^7 colonies. We used a diploid trp1- W303 strain as recipient, selecting on ScMin-2% glucose. We experienced sporadic culture contamination problems with environmental fungi; we found that use of canavanine (ScMin+can+lys) controlled this problem to a manageable level due to the *can1* canavanine-resistance mutation in W303. For control experiments we subcloned anti-GFP Nbs (22), also the anti-RBD nanobodies S1-1 and S1-23 (20), using standard cloning methods.

Surface nanobody induction

Stationary phase yeast were diluted 1:5 into 0.2% glucose-6%galactose in Sc-Min, grown overnight 30°C with rotation; supplemented with 10% volume fresh Sc-Min and an additional 3% galactose for another 24 hrs. These conditions gave strong induction of displayed nanobodies as detected using anti-GFP controls included in all experiments, so that we could detect surface nanobody by labeling cells with purified GFP. As described previously (56) we noted a substantial population of cells negative for GFP binding; plating experiments strongly suggested that these were due to plasmid loss events somehow induced by galactose incubation. Since these cells were plasmid-free they did not contribute to any downstream steps (outgrowth in selective medium or subsequent amplification of plasmid sequences), and therefore were inconsequential.

Yeast Affinity Capture

All antigens were conjugated to Thermo-Fisher Scientific (Waltham, MA) Dynabeads (M-270 epoxy 14301), following the manufacturer's protocol, with minor adaptations (57,58); GFP was made in-house (22) and the appropriate His-tagged SARS-CoV-2 spike antigens were obtained from Sino Biologicals (Chesterbrook, PA): Spike RBD, S1, or S2 ECD (original strain); Spike RBD L452, T478K (Delta); and B.1.1.529 (Omicron) Spike RBD. Yeast were vitally fluorescently labeled on their cell surfaces as previously described (40). We used GFP-Dynabeads and yeast expressing surface anti-GFP to establish conditions for binding and washing. The optimal binding buffer we discovered is described below. A 1 hr binding of yeast to beads with rotation at 30°C was followed by 4-5 washes with purification of bead-bound cells on a magnet using a Dynal MPC-6 magnetic stand, with samples kept at 2 cm from the magnet, 5 min binding per wash. All yeast affinity captures were performed in 1% BSA (Fraction V, protease-free; GoldBio (St. Louis, MO)), 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 pH to 7.4 with HCl / NaOH), 1% Tween-20 (Sigma Aldrich (St. Louis, MO)). For the epitope binning, affinity capture and subsequent library generation of the yeast was performed as described above, except that each 10 µl aliquot of the RBD-conjugated Dynabeads were pre-blocked with the addition of 20 µg of the appropriate nanobody or Ace2 in 1% BSA, 1 x PBS, 0.1% Tween-20 rotating for 1 hr at room temperature. Affinity capture with Miltenyi beads and subsequent fluorescence-activated cell sorting (FACS) were performed as described (34).

Sequencing of nanobody clones in the purified yeast library

After binding, beads with bound cells were transferred to ScMin-2% glucose and grown out for 14-48 hrs. Cells were pelleted, lysed with Zymolyase and DNA purified on Qiagen miniprep columns following manufacturer's procedures. The DNA prep was amplified with sequencing primers and sequenced at the Rockefeller Genomics facility using an Illumina MiSeq, PE250 (early experiments), PE300 (most experiments; better sequence quality due to longer overlap between the paired reads).

Nanobody cloning, expression and characterization
Cloning, expression and purification of the nanobodies, Surface Plasmon Resonance (SPR), Differential Scanning Fluorimetry (DSF) and SARS-CoV-2 Pseudovirus Neutralization Assays were performed as described (20).

Computational methods.

Nanobody sequences were obtained by paired-end sequencing (300 bp readlength) using Illumina MiSeq. Since each nanobody sequence was potentially represented by exactly one pair of reads, it was important to filter the data for quality. The computation was as follows: for positions covered only by one of the two paired-end reads, the quality score for that position was the one assigned by MiSeq. For positions covered by both of the paired-end reads (i.e., both strands sequenced), the base call was that for the higher-quality-scored position, and the final score was the sum of quality scores for the two reads if the base call was the same, and the higher minus the lower score if the base call was different. The overall nominal probability of having no error anywhere in the sequence was then computed as 1 – Product(-Q/10), product taken over all positions in the sequence, where Q is the final quality score at each position, and a cutoff of 0.9 applied. In addition, a similar calculation was applied to sequences approximately encoding CDR1, CDR2 and CDR3, with a cutoff of 0.95.

The CDR sequences were extracted from the complete nanobody sequence using the consensus FR region sequences from (41). Their consensus sequences were used to generate multiple alternate FR regions (usually with one or two substitutions each) and the best alignment to each FR (testing separately all of the candidate FR regions) was found. Sequences between FRs were assigned as CDRs 1,2,3. Subsequent computations were done using the 'CDR string' composed of the catenated CDR1,2,3 sequences. Due to minor FR variability there were approximately 1/3 as many CDR strings as full nanobody sequences. The data indicated strong concordance among nanobodies with the same CDR string, consistent with the known primacy of CDR sequence for binding specificity (see Introduction).

For all libraries, the number of sequences (nanobody or CDR string) was standardized to the total size of the library. Readcount was then adjusted to reads per million. The result was a table with rows corresponding to sequence and columns to standardized read count in a series of libraries. We created a custom viewer to compare readcount in various categories (e.g., in specified polygons in a 2-D graph; containing some CDR3 sequence; etc). The viewer employed log2(standardized readcount + 1); thus zero reads plotted at 0, 1 read at 1; higher readcounts plotted at approximately log2(readcount).

Crossover sequences (almost surely derived by PCR template-switching within the highly homologous FR regions) were identified as follows. Sequences (CDR strings) were ranked in order of abundance. The most abundant initiated a list of 'native' (non-crossover sequences). Subsequent (decreasing abundance) sequences were then examined for a good match in some CDRs to a sequence in the 'native' list combined with a bad match in other CDRs. Such cases

were assigned to a list of 'crossover' sequences; others (either distinct in all three CDRs, or similar in all three CDRs, to members of the native list) were appended to the native list.

All computations were carried out by MATLAB code, available upon request. Sequence logos and phylogenetic trees were calculated using built-in functions in the MATLAB Bioinformatics toolbox.

For the blocking experiment we used the standardized readcounts to calculate depletion/enrichment, based on #reads in blocked library / #reads in unblocked library. This was done after filtering out likely crossover products (see above).

DATA AVAILABILITY

All data generated or analyzed during this study are included in this published article (and its supplementary information), or are available from the corresponding authors on request.

MAIN FIGURES

Fig. 1. Design of Yeast-Based Nanobody Screen for anti-SARS-CoV-2 Spike Domains

Schematic of our yeast display based strategy for generating, identifying, and characterizing large, diverse repertoires of nanobodies that bind the spike protein of SARS-CoV-2. The highest quality nanobodies were assayed for their ability to neutralize SARS-CoV-2 pseudovirus. Figure adapted from(20); diagram of yeast display construct (bottom left) adapted from (34). Boxed inset shows 2.8 μm magnetic beads conjugated with S1 protein from SARS-CoV-2 Spike, after a binding reaction with yeast displaying an anti-S1 nanobody (top) or a nonspecific control (bottom), after non-binding yeast were washed away. Scale bar, 10 μm.

Fig. 2. Characterization of Competition Effects between Yeast Displaying Nanobodies of Different Defined Affinities.

A. Fluorescence microscopy of a yeast competition assay. Here, yeast displaying either LaG94-10 (Kd=2.9 pM) or LaG9 (Kd=3.9 nM) anti-GFP nanobodies were live surface-labeled with either Alexa594 (red) or Alexa350 (green)(40). These were mixed together in equal proportions (50:50:0, 'competitive' conditions; see text) with GFP-conjugated 2.8 μm magnetic beads (small green spheres), to which they bound (left panels, scale bar = 5 μm; see also Fig. 1). Right panel shows a low magnification field of the assay after several rounds of harvesting and washes, showing in this case the enrichment of the LaG94-10 displaying yeast; scale bar = 20 μm.

B. Plot of the coefficient of variation (CV) of yield for each yeast strain across the experiments in B against the Kd of their displayed nanobodies. As seen in the plot, the yields of yeast bearing low affinity nanobodies were highly variable specifically under competitive conditions. The yield of the highest-affinity Nb-bearing yeast, in contrast, were almost invariant because they always won the competition. Under non-competitive conditions, this differential was lost.

Fig. 3. Sequence Diversity of Nanobody Libraries.

Nanobody sequences from the unselected yeast library (from llama 7704; ~1.5×10^7 independent clones) were amplified and sequenced with Illumina Miseq and processed to minimize sequence errors, as described in Methods, yielding 1.2×10^6 distinct nanobody sequences. Framework regions (FRs) and complementarity-determining regions (CDRs) were determined in each encoded nanobody based on the alignment of (41). Unique sequences were determined and aligned by left justification of each region.

A. A seqlogo (44,59) was generated (MATLAB seqlogo command). High variability in the three CDR regions is evident.

B. Plot of proportion of non-consensus residues per residue across the library. At each position the sequence diversity was calculated (defined as the probability that two randomly chosen sequences are identical at the position (60). High variability in CDRs (red bars) is again observed, as well as lower but significant variation in the FRs (green bars).

C. Plot of sequence diversity per residue across the library (60). The library was sorted into CDR3 'families': groups of unique sequences (minimum family size 100) differing by no more than 20% in any of the seven regions (this criterion extracted about 15% of the unique sequences into 370 CDR3 families). Because this criterion results in high similarity of CDRs within a CDR3 family, it is very likely that the sequences in each CDR3 family derive from a unique VDJ recombination with subsequent somatic hypermutation. For each CDR3 family the sequence diversity was calculated, and the aggregate average diversity graphed as in B. As expected, the diversity within CDR3 families was significantly reduced compared to the whole library (note difference in y-axis scale), and the differential between FRs and CDRs largely lost.

Fig. 4. Validation of the Yeast Display Method with Biochemically Identified Nanobodies.

A. Behavior of the yeast clones carrying CDRs matching the mass spectrometry positive nanobodies (i.e., 'MS positives') in the Spike domain affinity selection assay. The read counts in the unselected and selected libraries for all these sequences are plotted on top of the plot of the overall results from the entire llama 7704 nanobody cDNA display library (gray points), as plots of the log2 of these values + 1, as shown in Fig. S1. 1× selection: one round of binding. 2×: selection: yeast from the 1× selection were grown out, expression of the nanobody fusion re-induced, and the yeast were bound to the same antigen (see Results, Methods); screens were for binding to either the S1 domain (S1 selection), RBD domain (RBD selection), or S2 domain (S2 selection) of Spike. In green are nanobodies that were shown previously to bind to S1 but not RBD; in blue are those that bound to both RBD and S1; in red are binders to S2. Note that all these nanobodies bound with precisely the specificity based on prior biochemical characterization (20) with the exception of a small number that failed to bind in any of the selections. TheThese 'MS positives' generally bound well in both the 1× and 2× selections. These results support the specificity and comprehensiveness of the yeast display procedure.

B. Correlation of yeast display binding with the affinity of their displayed nanobodies, for the MS-positives plotted in part A. Plot of enrichment (log2(bound)-log2(unselected) versus log10(nanobody Kd). Top: 1× selection. Bottom: 2× selection. Pearson R values and associated P values by t-test are shown. The figure shows a noisy but significant correlation between binding affinity and enrichment by yeast display binding, especially for the more highly competitive 2× binding assay (see text).

Fig. 5. Screening for New Families of Anti-Spike Nanobodies.

A. The read counts in the unselected and 1× and 2× selected libraries screened against either the S1, RBD or S2 domains of Spike from the entire llama 7704 nanobody cDNA display library (gray points) are plotted, as the log2 of these values + 1, as shown in Fig. 4. Sequences displaying different specificities were identified and selected from the graphical display as in Fig. 4, except here nanobodies specific for only the RBD subdomain of S1 were separately labeled from those that recognized the non-RBD portions of S1. Selection was by a differential polygon function allowing multiple criteria. For example, sequences in an enriched polygon for S1 but not for RBD defines nanobodies binding to the non-RBD subdomain of S1 (also see Introduction). Green: S1 non-RBD; blue: RBD; red: S2. The darker colors show the initial CDR families taken from the ~200 highest read counts sets; the lighter colors represent an extension of these CDR families to a broader set allowing up to 20% changes in any of the three CDR sequences, and above a minimum read count. Note that all these CDR families are bound with consistent specificity, indicating that in general, sequence variations within the families (see Figure 3) do not strongly affect function.

B. To evaluate sequence diversity within these functional classes, we employed the MATLAB command 'phytree' to construct a neighbor-joining tree for the ~200 abundant CDR sequences in each class (i.e., using the sequences identified by the dark green, blue or red points in panel A). A wide diversity of sequences was observed, which sorted out into a more limited set of CDR families of related sequences. Families with generally larger representation are labeled with Roman numerals; black dots indicate sequences that were expressed and characterized as recombinant monomers (see below); these were chosen from the tree to sample broadly across the sequence space, while minimizing repetitive sampling of families with clones already characterized from the mass-spectrometrically identified nanobodies (20).

C. Sequence conservation within the CDR families indicated in (B) was evaluated using the MATLAB seqlogo command (44) applied to the CDR strings of the selected CDR families; shown are SeqLogos of (left to right in each SeqLogo) CDR1, CDR2 and CDR3.

Fig. 6. Testing the Standard and Shuffled Libraries against RBD variants Delta and Omicron.

A. Testing the standard library against RBD variants Delta and Omicron. Dynabeads conjugated with RBD from the original SARS-CoV-2 and from the Delta and Omicron variants were employed for affinity purification of yeast display clones from the llama 7704 library. Two rounds of selection were carried out, as in Figs. 4 and 5. A clear overall reduction in binding was observed, though many clones still bound well to both variants. Plotted on top of the overall library result (gray points) are sequence families illustrating some of the main patterns of response to the variants are indicated as colored dots, with each color corresponding to the indicated family, defined here by their core CDR3 amino acid sequence. These families contain considerable sequence diversity within them (Fig. 5), and represent much larger numbers of many other unrelated clones (often with lower representation in the library) that exhibit similar behaviors.
B. Testing a shuffled anti-RBD library for CDR autonomy and rescued activity in recombinants. (Top). Binding of the shuffled library to RBD. One round of purification was

carried out, followed by sequencing. Results were analyzed based on the fate of CDR3 'groups' (where within a group, a given CDR3 bound to a high diversity of CDR1,2 recombinants). Different behaviors were observed; three are illustrated with colored dots (same color scheme as Fig. 7). The 'NAAAW' group (green) bound well to original and variant RBDs, essentially independent of the recombinant CDRs 1 and 2 it was attached to, suggesting strong 'CDR3 dominance' of the 'NAAAW' CDR3. Similarly, the 'IIDDY' group exhibited essentially similar behavior when shuffled as when combined with its native CDR1,2 (Fig. 7): strong binding to original and Omicron, but clearly weaker binding to Delta. The 'YERLAWD' recombinant group bound comparably to all variants, in contrast to its behavior with its native CDRs 1 and 2, which rendered it unable to bind Delta or Omicron.

(Bottom). Effectiveness of the shuffle is shown by extracting all sequences bearing a specific CDR3 and examining sequence diversity of CDRs 1 and 2. Sequence logos show that highly diverse CDRs 1 and 2 are observed joined to each of three different CDR3's, and the pattern of CDR1,2 diversity attached to the CDR3's was essentially the same.

Fig. 7. Biophysical and Neutralization Properties of the Nanobodies

Thirty yeast display nanobodies targeting the S1-RBD, S1 non-RBD, and S2 portions of spike were functionally tested for neutralization of lentivirus pseudotyped with various SARS-CoV spikes and their biophysical properties characterized. (Red text indicates YERLAWD family members, and blue text indicates NAAAW family members, used in the crossover 'rescue' studies; see text). (A) Neutralization data against Original, Delta and Omicron. (B) Kd measurements of 21 of these nanobodies were determined using SPR and affinities plotted. Nanobodies were tested against Original, Delta and Omicron recombinant S1 or RBD. S1 non-RBD nanobodies were not tested against Omicron. Kd measurements for three of the 'rescue' constructs plotted at right. (C) The Tm measurements of the nanobodies in (B) were determined using DSF and plotted. CoV2-YD-6 and CoV2-YD-38 (highlighted in gray) resulted in two distinct melting peaks. Open circles indicate less proportion of this species in the sample. Tm measurements for three of the 'rescue' constructs plotted at right.

Fig. 8. Epitope Binning by Yeast Display.

Dynabeads conjugated with RBD were blocked with monomer nanobodies representing the 7 epitope classes defined previously (20); with the soluble extracellular domain of the RBD target Ace2; or left unblocked. The 2×-RBD-selected library from llama 7704 (Figs. 4, 5, S1) was bound to these beads. The bound VHHs were sequenced, and enrichment/depletion upon blocking for each 'CDR string' (catenated CDR1/CDR2/CDR3) was calculated as log2(readcount with blocked beads/readcount with unblocked beads). The sequences were filtered to remove PCR crossover artifacts (Methods). Left: The resulting matrix of ~100,000 sequences X 8 blocking agents was filtered for readcount (at least 100 reads combining all blocking experiments) and clustered using the MATLAB hierarchical clustering algorithm; scale bar on left indicates log2 enrichment/depletion (above). Right: The location of sequences containing the indicated CDR3, or close relatives, was determined. Note that the hierarchical clustering was sequence-blind, so the clustering of related sequences implies similar blocking behavior across the family. Bottom: Representative nanobodies (yellow ribbon) binding to their respective epitopes

(blue surface patch), as defined previously (20), are depicted on the receptor binding domain (RBD) of SARS-CoV-2 spike (PDB ID: 6M0J).

Supplementary Figures

Fig. S1. Binding of nanobodies expressed in yeast to the different recombinant S1, RBD and S2 domains of SARS-Cov2 Spike protein (labeled S1, RBD, and S2 selection respectively) (20) conjugated to Dynabeads.
Binding and washing were as described in Methods. Nanobody sequences were amplified and sequenced as in Fig. 3. The CDR regions were extracted from each sequence, and concatenated in a 'CDR string'. The read counts for each unique CDR string in the unselected library and the selected libraries were compared and, after standardization to reads per million nanobody sequences, plotted (gray points). Because of the wide spread of values, we plot the log2 of these values + 1 (the addition of 1 allows plotting of zero reads as the value 0). For all three antigens we observe two lines of points, one above the x=y line and one below. Both lines trend upwards with increasing read count in the unselected library. We take the upper line to represent specific binding (enriched; labeled with red oval as "specific binders"), and the lower line nonspecific binding (depleted; labeled "nonspecific binders"); the blue line (x=y) thus represents neither enrichment nor depletion. Both lines increase with a slope of one, probably simply because there are more yeast bearing that sequence due to differential abundance in the llama. The vertical displacement of the specific from the non-specific line represents the overall enrichment of specific over non-specific binding in our conditions, approximately 1000-fold. This is in agreement with the differential binding we achieved comparing yeast with a control surface nanobody to yeast with a high-affinity anti-GFP nanobody to Dynabeads bearing GFP (Figure 2). 1× selection: one round of binding. 2×: selection: yeast from the 1× selection were grown out, expression of the nanobody fusion re-induced, and the yeast were bound to the same antigen (see Methods). The main effect of this was to reduce the population of the lower (nonspecific) line.

Fig. S2. Validation of yeast display method with MS-positives from another llama.
In (20) we searched for nanobodies in two llamas, 7704 ("Rocky") and 5094 ("Marley"). High-affinity binders were recovered from llama 5094, though apparently fewer than from 7704, reflecting the results of(20). We carried out the same assay for these nanobodies and plotted the results as in Fig. 4A, with the exception that to get sufficient representation we accepted CDR strings with up to 20% sequence variation in each CDR; each different color of the dots plotted represents a member of a family of a given mass spectrometry-positive nanobody. We observed the same high specificity and recovery in yeast display for the 5094 clones as for the 7704 clones. Note that in contrast to 7704, 2X binding was required for a clean result in some cases; this is likely due to the overall lower anti-Spike nanobody content in 5094 noted previously (20), which increases the level of non-specific binding in the 1×.

Fig. S3. Comparison of different methods for screening the display libraries.
The entire llama 7704 nanobody cDNA display library was screened against the RBD domain of Spike, and plotted as in Fig. S1 (gray points). Two methods of screening were employed. The first was as described in Fig. 1 and Methods, using two rounds of Dynabead selection. The second was following a published procedure with the first selection utilizing Miltenyi magnetic beads and biotin purification, which was followed by

a second selection utilizing a FACS sorting protocol with Alexa-labeled RBD (34) (Methods).

A. All sequences recovered are plotted as in Fig. S1.
B. Families of MS-positive sequences are plotted on top of the overall graph, as in Fig. S2.

Fig. S4.Testing CDR families of mass-spectrometry positives against RBD variants Delta and Omicron.
The data are plotted as in Figure 7, but plotting CDR families of the mass-spectrometry-positive clones; each different color of the dots plotted represents a different mass spectrometry-positive nanobody.

Fig. S5. Diagram of the splice-overlap extension method for recombining CDRs.
Degenerate oligos priming in both directions from two highly conserved sequences within framework regions (FRs) 2 and 3, and end oligos tagged for recombination onto the yeast expression vector, were used to amplify fragments containing CDRs 1, 2 and 3 as well as flanking framework regions. The template used was a pool after 1 round of selection on RBD. The result of the sequential PCR reactions indicated is a random mix-and-match of the three CDRs and flanking framework regions.

Fig. S6. Epitope Binning Tests.
Dynabeads conjugated with RBD were pre-blocked with saturating amounts of either monomer nanobodies S1-1 or S1-23, with unblocked Dynabeads as control. Yeast expressing the anti-RBD nanobodies S1-1, or S1-23, or S1-RBD-38 (each belonging to different epitope classes as defined by (20)), or the anti-GFP nanobody LaG94-10, were bound to these beads. After binding and washing, we quantified recovery of yeast by colony counts of serial dilutions on yeast medium. S1-1-expressing and S1-23-expressing yeast were blocked from binding by the cognate but not the non-cognate nanobody block.

Fig. S7. Epitope Sub Clusters of Families.
Left: same as Figure 8 left, for reference. The indicated subclusters were extracted using the MATLAB clustergram tool and CDR sequences recovered (right).

Fig. S8. Crossover Epitope Binning.
The sequences clustered in Fig. 8 were curated computationally to eliminate a population of PCR crossover events (see Methods) - these are essentially similar to the ‘shuffled’ products made intentionally (Fig. 6). We analyzed these crossover products separately (left), and observed intriguing behavior in some cases. For example, while blockage by #6 and #7 was mutually exclusive in the native sequences (Fig. 8), crossover products with CDR1 and 2 derived from the ‘YERLAWD’ native sequence and CDR3 from the ‘NAAAW’ native family were blocked by *both* #6 and #7; the converse recombinants (CDR1/2 from ‘NAAAW’ and CDR3 from ‘YERLAWD’) behaved similarly (these sequences provide almost the entire signal in the boxed region in the figure). Numerous other examples of similar apparent ‘hybrid’ epitope classes were observed among the crossover recombinants. Biochemical analysis would be required to determine if such behavior represents a bona fide two-part epitope with CDR1/2 binding at one place and CDR3 binding at another; however, at present we cannot exclude that the ‘hybrid epitope classes’ are specific to the potentially multi-point attachment of polyvalent yeast to a polyvalent binding surface.

REFERENCES

1. Fauci, A. S. (2021) The story behind COVID-19 vaccines. *Science* **372**, 109
2. Ball, P. (2021) The lightning-fast quest for COVID vaccines - and what it means for other diseases. *Nature* **589**, 16-18
3. Adam, D. (2022) The pandemic's true death toll: millions more than official counts. *Nature* **601**, 312-315
4. Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Esteban, O.-O., Hasell, J., Macdonal, B., Beltekian, D., and Roser, M. (2020) Coronavirus Pandemic (COVID-19). OurWorldInData.org
5. Shaman, J., and Galanti, M. (2020) Will SARS-CoV-2 become endemic? *Science* **370**, 527-529
6. Antia, R., and Halloran, M. E. (2021) Transition to endemicity: Understanding COVID-19. *Immunity* **54**, 2172-2176
7. Letko, M., Marzi, A., and Munster, V. (2020) Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* **5**, 562-569
8. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020) Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330-333
9. Hsieh, C. L., Goldsmith, J. A., Schaub, J. M., DiVenere, A. M., Kuo, H. C., Javanmardi, K., Le, K. C., Wrapp, D., Lee, A. G., Liu, Y., Chou, C. W., Byrne, P. O., Hjorth, C. K., Johnson, N. V., Ludes-Meyers, J., Nguyen, A. W., Park, J., Wang, N., Amengor, D., Lavinder, J. J., Ippolito, G. C., Maynard, J. A., Finkelstein, I. J., and McLellan, J. S. (2020) Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501-1505
10. Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., Zheng, X. S., Zhao, K., Chen, Q. J., Deng, F., Liu, L. L., Yan, B., Zhan, F. X., Wang, Y. Y., Xiao, G. F., and Shi, Z. L. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273
11. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B. S., and McLellan, J. S. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263
12. Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292 e286
13. Chmielewska, A. M., Czarnota, A., Bienkowska-Szewczyk, K., and Grzyb, K. (2021) Immune response against SARS-CoV-2 variants: the role of neutralization assays. *NPJ Vaccines* **6**, 142
14. Planas, D., Veyer, D., Baidaliuk, A., Staropoli, I., Guivel-Benhassine, F., Rajah, M. M., Planchais, C., Porrot, F., Robillard, N., Puech, J., Prot, M., Gallais, F., Gantner, P., Velay, A., Le Guen, J., Kassis-Chikhani, N., Edriss, D., Belec, L., Seve, A., Courtellemont, L., Pere, H., Hocqueloux, L., Fafi-Kremer, S., Prazuck, T., Mouquet, H., Bruel, T., Simon-Loriere, E., Rey, F. A., and Schwartz, O. (2021) Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276-280

15. Wilhelm, A., Widera, M., Grikscheit, K., Toptan, T., Schenk, B., Pallas, C., Metzler, M., Kohmer, N., Hoehl, S., Helfritz, F. A., Wolf, T., Goetsch, U., and Ciesek, S. (2021) Reduced Neutralization of SARS-CoV-2 Omicron Variant by Vaccine Sera and Monoclonal Antibodies. *medRxiv*
16. Cameroni, E., Bowen, J. E., Rosen, L. E., Saliba, C., Zepeda, S. K., Culap, K., Pinto, D., VanBlargan, L. A., De Marco, A., di Iulio, J., Zatta, F., Kaiser, H., Noack, J., Farhat, N., Czudnochowski, N., Havenar-Daughton, C., Sprouse, K. R., Dillen, J. R., Powell, A. E., Chen, A., Maher, C., Yin, L., Sun, D., Soriaga, L., Bassi, J., Silacci-Fregni, C., Gustafsson, C., Franko, N. M., Logue, J., Iqbal, N. T., Mazzitelli, I., Geffner, J., Grifantini, R., Chu, H., Gori, A., Riva, A., Giannini, O., Ceschi, A., Ferrari, P., Cippa, P. E., Franzetti-Pellanda, A., Garzoni, C., Halfmann, P. J., Kawaoka, Y., Hebner, C., Purcell, L. A., Piccoli, L., Pizzuto, M. S., Walls, A. C., Diamond, M. S., Telenti, A., Virgin, H. W., Lanzavecchia, A., Snell, G., Veesler, D., and Corti, D. (2022) Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature* **602**, 664-670
17. Farid, S. S., Baron, M., Stamatis, C., Nie, W., and Coffman, J. (2020) Benchmarking biopharmaceutical process development and manufacturing cost contributions to R&D. *MAbs* **12**, 1754999
18. Muyldermans, S. (2021) A guide to: generation and design of nanobodies. *FEBS J* **288**, 2084-2102
19. Ciccarese, S., Burger, P. A., Ciani, E., Castelli, V., Linguiti, G., Plasil, M., Massari, S., Horin, P., and Antonacci, R. (2019) The Camel Adaptive Immune Receptors Repertoire as a Singular Example of Structural and Functional Genomics. *Front Genet* **10**, 997
20. Mast, F. D., Fridy, P. C., Ketaren, N. E., Wang, J., Jacobs, E. Y., Olivier, J. P., Sanyal, T., Molloy, K. R., Schmidt, F., Rutkowska, M., Weisblum, Y., Rich, L. M., Vanderwall, E. R., Dambrauskas, N., Vigdorovich, V., Keegan, S., Jiler, J. B., Stein, M. E., Olinares, P. D. B., Herlands, L., Hatziioannou, T., Sather, D. N., Debley, J. S., Fenyo, D., Sali, A., Bieniasz, P. D., Aitchison, J. D., Chait, B. T., and Rout, M. P. (2021) Highly synergistic combinations of nanobodies that target SARS-CoV-2 and are resistant to escape. *Elife* **10**
21. Muyldermans, S. (2013) Nanobodies: natural single-domain antibodies. *Annu Rev Biochem* **82**, 775-797
22. Fridy, P. C., Li, Y., Keegan, S., Thompson, M. K., Nudelman, I., Scheid, J. F., Oeffinger, M., Nussenzweig, M. C., Fenyo, D., Chait, B. T., and Rout, M. P. (2014) A robust pipeline for rapid production of versatile nanobody repertoires. *Nat Methods* **11**, 1253-1260
23. Xu, J., Xu, K., Jung, S., Conte, A., Lieberman, J., Muecksch, F., Lorenzi, J. C. C., Park, S., Schmidt, F., Wang, Z., Huang, Y., Luo, Y., Nair, M. S., Wang, P., Schulz, J. E., Tessarollo, L., Bylund, T., Chuang, G. Y., Olia, A. S., Stephens, T., Teng, I. T., Tsybovsky, Y., Zhou, T., Munster, V., Ho, D. D., Hatziioannou, T., Bieniasz, P. D., Nussenzweig, M. C., Kwong, P. D., and Casellas, R. (2021) Nanobodies from camelid mice and llamas neutralize SARS-CoV-2 variants. *Nature* **595**, 278-282
24. Tang, Q., Owens, R. J., and Naismith, J. H. (2021) Structural Biology of Nanobodies against the Spike Protein of SARS-CoV-2. *Viruses* **13**
25. Liu, H., Yuan, M., Huang, D., Bangaru, S., Zhao, F., Lee, C. D., Peng, L., Barman, S., Zhu, X., Nemazee, D., Burton, D. R., van Gils, M. J., Sanders, R. W., Kornau, H. C., Reincke, S. M., Pruss, H., Kreye, J., Wu, N. C., Ward, A. B., and Wilson, I. A. (2021) A combination of cross-neutralizing antibodies synergizes to prevent SARS-CoV-2 and SARS-CoV pseudovirus infection. *Cell Host Microbe* **29**, 806-818 e806

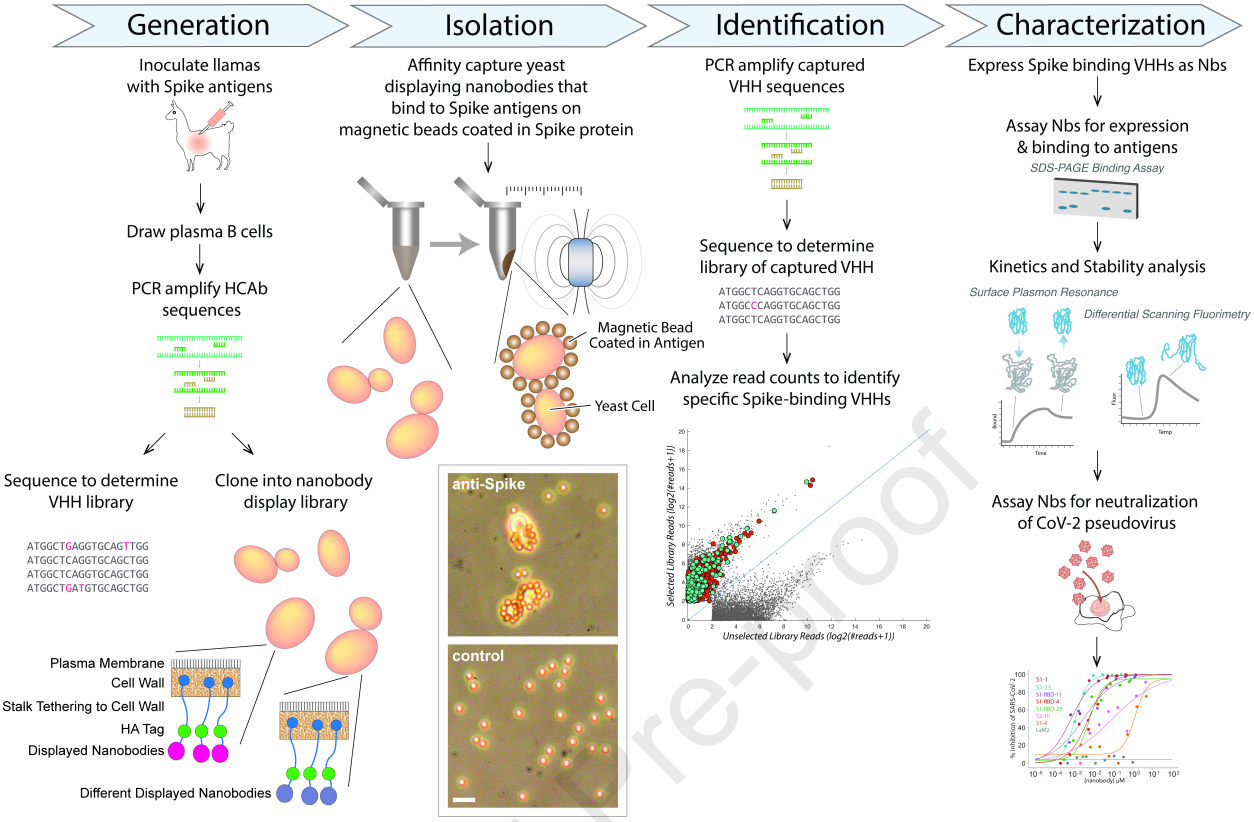
26. Dong, J., Huang, B., Jia, Z., Wang, B., Gallolu Kankanamalage, S., Titong, A., and Liu, Y. (2020) Development of multi-specific humanized llama antibodies blocking SARS-CoV-2/ACE2 interaction with high affinity and avidity. *Emerg Microbes Infect* **9**, 1034-1036
27. Van Heeke, G., Allosery, K., De Brabandere, V., De Smedt, T., Detalle, L., and de Fougerolles, A. (2017) Nanobodies(R) as inhaled biotherapeutics for lung diseases. *Pharmacol Ther* **169**, 47-56
28. Nambulli, S., Xiang, Y., Tilston-Lunel, N. L., Rennick, L. J., Sang, Z., Klimstra, W. B., Reed, D. S., Crossland, N. A., Shi, Y., and Duprex, W. P. (2021) Inhalable Nanobody (PiN-21) prevents and treats SARS-CoV-2 infections in Syrian hamsters at ultra-low doses. *Sci Adv* **7**
29. Revets, H., De Baetselier, P., and Muyldermans, S. (2005) Nanobodies as novel agents for cancer therapy. *Expert Opin Biol Ther* **5**, 111-124
30. Jovcevska, I., and Muyldermans, S. (2020) The Therapeutic Potential of Nanobodies. *BioDrugs* **34**, 11-26
31. Bannas, P., Hambach, J., and Koch-Nolte, F. (2017) Nanobodies and Nanobody-Based Human Heavy Chain Antibodies As Antitumor Therapeutics. *Front Immunol* **8**, 1603
32. Peeling, R. W., and McNerney, R. (2014) Emerging technologies in point-of-care molecular diagnostics for resource-limited settings. *Expert Rev Mol Diagn* **14**, 525-534
33. Thompson, M. K., Fridy, P. C., Keegan, S., Chait, B. T., Fenyo, D., and Rout, M. P. (2016) Optimizing selection of large animals for antibody production by screening immune response to standard vaccines. *J Immunol Methods* **430**, 56-60
34. McMahon, C., Baier, A. S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J. X., Erlandson, S. C., Hilger, D., Rasmussen, S. G. F., Ring, A. M., Manglik, A., and Kruse, A. C. (2018) Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat Struct Mol Biol* **25**, 289-296
35. Romao, E., Morales-Yanez, F., Hu, Y., Crauwels, M., De Pauw, P., Hassanzadeh, G. G., Devoogdt, N., Ackaert, C., Vincke, C., and Muyldermans, S. (2016) Identification of Useful Nanobodies by Phage Display of Immune Single Domain Libraries Derived from Camelid Heavy Chain Antibodies. *Curr Pharm Des* **22**, 6500-6518
36. Romao, E., Poignavent, V., Vincke, C., Ritzenthaler, C., Muyldermans, S., and Monsion, B. (2018) Construction of High-Quality Camel Immune Antibody Libraries. *Methods Mol Biol* **1701**, 169-187
37. Roth, L., Krah, S., Klemm, J., Gunther, R., Toleikis, L., Busch, M., Becker, S., and Zielonka, S. (2020) Isolation of Antigen-Specific VHH Single-Domain Antibodies by Combining Animal Immunization with Yeast Surface Display. *Methods Mol Biol* **2070**, 173-189
38. Schoof, M., Faust, B., Saunders, R. A., Sangwan, S., Rezelj, V., Hoppe, N., Boone, M., Billesbolle, C. B., Puchades, C., Azumaya, C. M., Kratochvil, H. T., Zimanyi, M., Deshpande, I., Liang, J., Dickinson, S., Nguyen, H. C., Chio, C. M., Merz, G. E., Thompson, M. C., Diwanji, D., Schaefer, K., Anand, A. A., Dobzinski, N., Zha, B. S., Simoneau, C. R., Leon, K., White, K. M., Chio, U. S., Gupta, M., Jin, M., Li, F., Liu, Y., Zhang, K., Bulkley, D., Sun, M., Smith, A. M., Rizo, A. N., Moss, F., Brilot, A. F., Pourmal, S., Trenker, R., Pospiech, T., Gupta, S., Barsi-Rhyne, B., Belyy, V., Barile-Hill, A. W., Nock, S., Liu, Y., Krogan, N. J., Ralston, C. Y., Swaney, D. L., Garcia-Sastre, A., Ott, M., Vignuzzi, M., Consortium, Q. S.

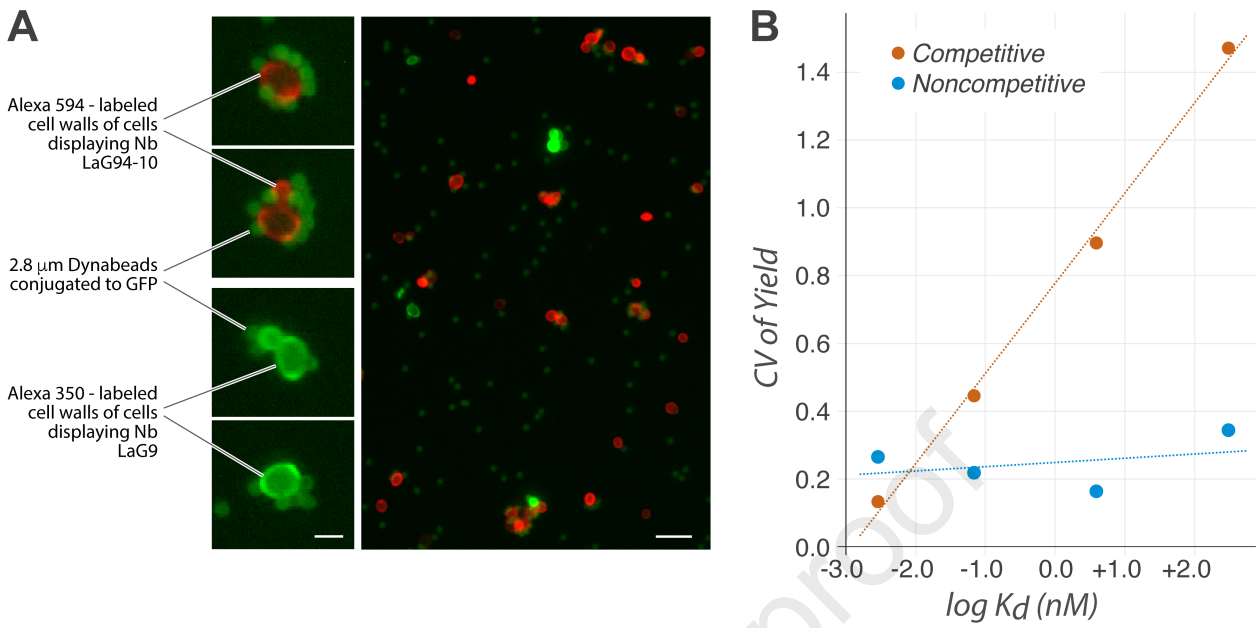
B., Walter, P., and Manglik, A. (2020) An ultrapotent synthetic nanobody neutralizes SARS-CoV-2 by stabilizing inactive Spike. *Science* **370**, 1473-1479

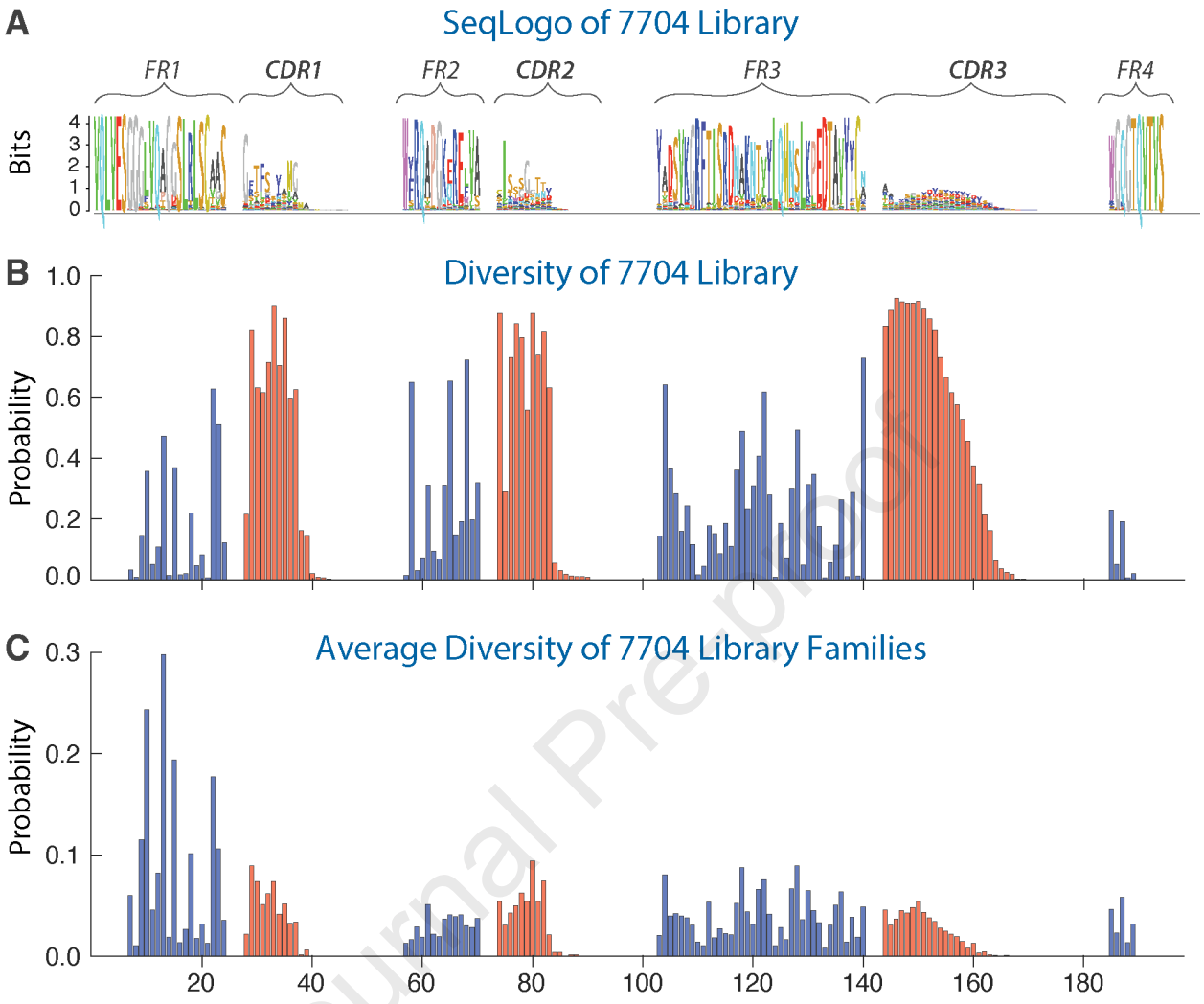
39. Uchanski, T., Zogg, T., Yin, J., Yuan, D., Wohlkonig, A., Fischer, B., Rosenbaum, D. M., Kobilka, B. K., Pardon, E., and Steyaert, J. (2019) An improved yeast surface display platform for the screening of nanobody immune libraries. *Sci Rep* **9**, 382
40. Timney, B. L., Tetenbaum-Novatt, J., Agate, D. S., Williams, R., Zhang, W., Chait, B. T., and Rout, M. P. (2006) Simple kinetic relationships and nonspecific competition govern nuclear import rates in vivo. *J Cell Biol* **175**, 579-593
41. Mitchell, L. S., and Colwell, L. J. (2018) Comparative analysis of nanobody sequence and structure data. *Proteins* **86**, 697-706
42. Zavrtanik, U., Lukan, J., Loris, R., Lah, J., and Hadzi, S. (2018) Structural Basis of Epitope Recognition by Heavy-Chain Camelid Antibodies. *J Mol Biol* **430**, 4369-4386
43. Fridy, P. C., Thompson, M. K., Ketaren, N. E., and Rout, M. P. (2015) Engineered high-affinity nanobodies recognizing staphylococcal Protein A and suitable for native isolation of protein complexes. *Anal Biochem* **477**, 92-94
44. Schneider, T. D., and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18**, 6097-6100
45. Greaney, A. J., Starr, T. N., Gilchuk, P., Zost, S. J., Binshtein, E., Loes, A. N., Hilton, S. K., Huddleston, J., Eguia, R., Crawford, K. H. D., Dingens, A. S., Nargi, R. S., Sutton, R. E., Suryadevara, N., Rothlauf, P. W., Liu, Z., Whelan, S. P. J., Carnahan, R. H., Crowe, J. E., Jr., and Bloom, J. D. (2021) Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host Microbe* **29**, 44-57 e49
46. Wang, L., Shi, W., Chappell, J. D., Joyce, M. G., Zhang, Y., Kanekiyo, M., Becker, M. M., van Doremalen, N., Fischer, R., Wang, N., Corbett, K. S., Choe, M., Mason, R. D., Van Galen, J. G., Zhou, T., Saunders, K. O., Tatti, K. M., Haynes, L. M., Kwong, P. D., Modjarrad, K., Kong, W. P., McLellan, J. S., Denison, M. R., Munster, V. J., Mascola, J. R., and Graham, B. S. (2018) Importance of Neutralizing Monoclonal Antibodies Targeting Multiple Antigenic Sites on the Middle East Respiratory Syndrome Coronavirus Spike Glycoprotein To Avoid Neutralization Escape. *J Virol* **92**
47. Xiang, Y., Nambulli, S., Xiao, Z., Liu, H., Sang, Z., Duprex, W. P., Schneidman-Duhovny, D., Zhang, C., and Shi, Y. (2020) Versatile and multivalent nanobodies efficiently neutralize SARS-CoV-2. *Science* **370**, 1479-1484
48. Huo, J., Le Bas, A., Ruza, R. R., Duyvesteyn, H. M. E., Mikolajek, H., Malinauskas, T., Tan, T. K., Rijal, P., Dumoux, M., Ward, P. N., Ren, J., Zhou, D., Harrison, P. J., Weckener, M., Clare, D. K., Vogirala, V. K., Radecke, J., Moynie, L., Zhao, Y., Gilbert-Jaramillo, J., Knight, M. L., Tree, J. A., Buttigieg, K. R., Coombes, N., Elmore, M. J., Carroll, M. W., Carrique, L., Shah, P. N. M., James, W., Townsend, A. R., Stuart, D. I., Owens, R. J., and Naismith, J. H. (2020) Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. *Nat Struct Mol Biol* **27**, 846-854
49. Pymm, P., Adair, A., Chan, L. J., Cooney, J. P., Mordant, F. L., Allison, C. C., Lopez, E., Haycroft, E. R., O'Neill, M. T., Tan, L. L., Dietrich, M. H., Drew, D., Doerflinger, M., Dengler, M. A., Scott, N. E., Wheatley, A. K., Gherardin, N. A., Venugopal, H., Cromer, D., Davenport, M. P., Pickering, R., Godfrey, D. I., Purcell, D. F. J., Kent, S. J., Chung, A. W., Subbarao, K., Pellegrini, M., Glukhova, A., and Tham, W. H. (2021) Nanobody cocktails

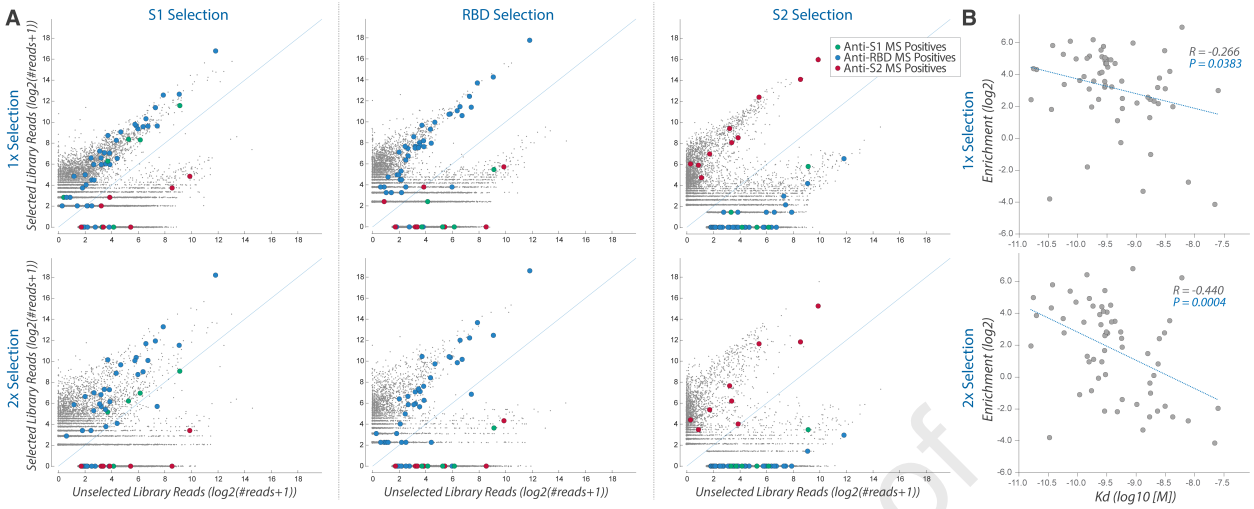
potently neutralize SARS-CoV-2 D614G N501Y variant and protect mice. *Proc Natl Acad Sci U S A* **118**

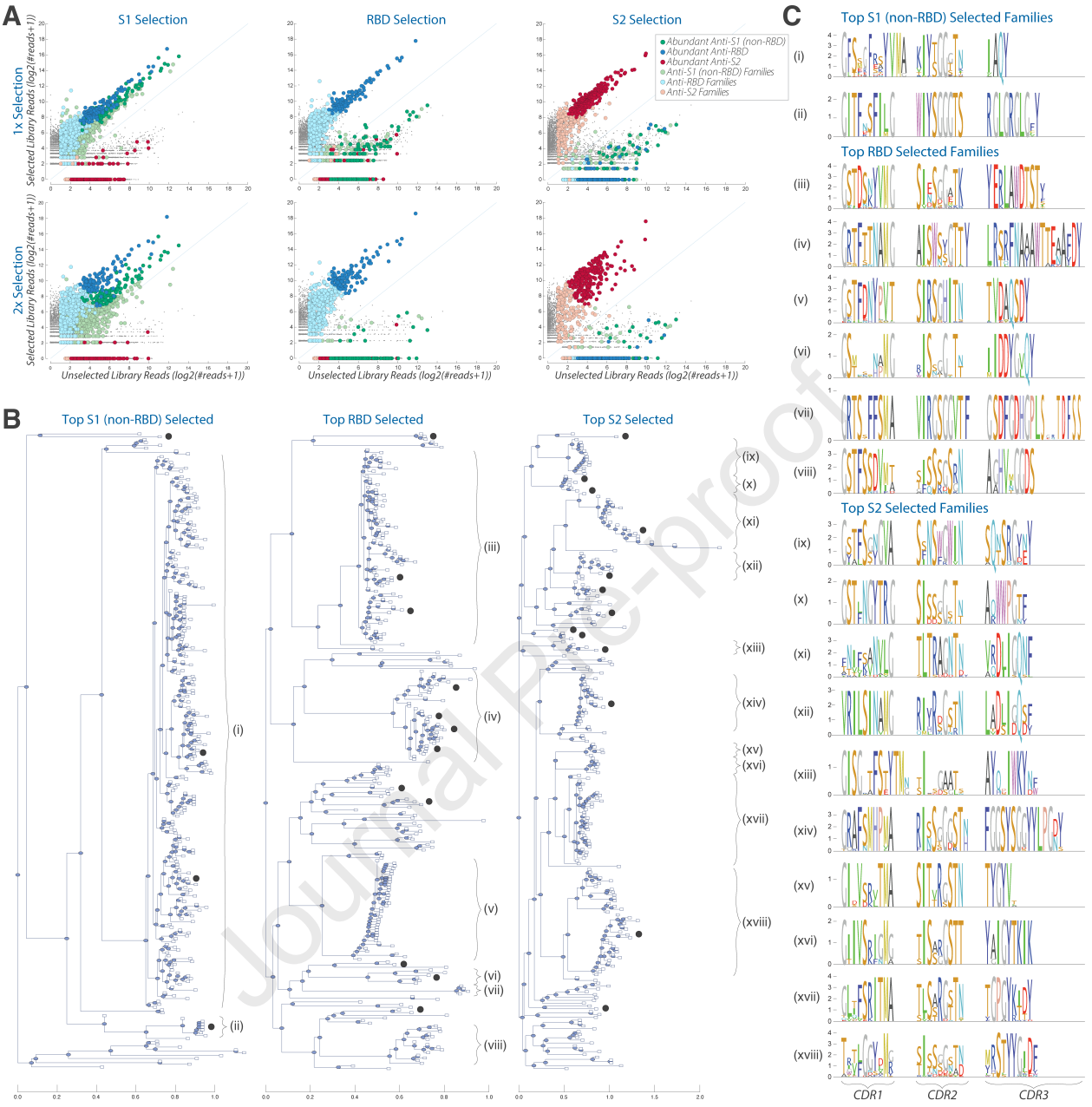
50. Starr, T. N., Greaney, A. J., Hannon, W. W., Loes, A. N., Hauser, K., Dillen, J. R., Ferri, E., Farrell, A. G., Dadonaite, B., McCallum, M., Matreyek, K. A., Corti, D., Veesler, D., Snell, G., and Bloom, J. D. (2022) Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. *Science* **377**, 420-424
51. Greaney, A. J., Starr, T. N., Barnes, C. O., Weisblum, Y., Schmidt, F., Caskey, M., Gaebler, C., Cho, A., Agudelo, M., Finkin, S., Wang, Z., Poston, D., Muecksch, F., Hatziioannou, T., Bieniasz, P. D., Robbiani, D. F., Nussenzweig, M. C., Bjorkman, P. J., and Bloom, J. D. (2021) Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nat Commun* **12**, 4196
52. Sheedy, C., Yau, K. Y., Hirama, T., MacKenzie, C. R., and Hall, J. C. (2006) Selection, characterization, and CDR shuffling of naive llama single-domain antibodies selected against auxin and their cross-reactivity with auxinic herbicides from four chemical families. *J Agric Food Chem* **54**, 3668-3678
53. Zupancic, J. M., Desai, A. A., and Tessier, P. M. (2022) Facile isolation of high-affinity nanobodies from synthetic libraries using CDR-swapping mutagenesis. *STAR Protoc* **3**, 101101
54. Tsukahara, N., Murakami, A., Motohashi, M., Nakayama, H., Kondo, Y., Ito, Y., Azuma, T., and Kishimoto, H. (2022) An alpaca single-domain antibody (VHH) phage display library constructed by CDR shuffling provided high-affinity VHHs against desired protein antigens. *Int Immunol* **34**, 421-434
55. Benatuil, L., Perez, J. M., Belk, J., and Hsieh, C. M. (2010) An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel* **23**, 155-159
56. Zupancic, J. M., Desai, A. A., Schardt, J. S., Pornnoppadol, G., Makowski, E. K., Smith, M. D., Kennedy, A. A., Garcia de Mattos Barbosa, M., Cascalho, M., Lanigan, T. M., Tai, A. W., and Tessier, P. M. (2021) Directed evolution of potent neutralizing nanobodies against SARS-CoV-2 using CDR-swapping mutagenesis. *Cell Chem Biol* **28**, 1379-1388 e1377
57. Obado, S. O., Field, M. C., Chait, B. T., and Rout, M. P. (2016) High-Efficiency Isolation of Nuclear Envelope Protein Complexes from Trypanosomes. *Methods Mol Biol* **1411**, 67-80
58. Kim, S. J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B. D., Hogan, J. A., Upla, P., Chemmama, I. E., Pellarin, R., Echeverria, I., Shivaraju, M., Chaudhury, A. S., Wang, J., Williams, R., Unruh, J. R., Greenberg, C. H., Jacobs, E. Y., Yu, Z., de la Cruz, M. J., Mironska, R., Stokes, D. L., Aitchison, J. D., Jarrold, M. F., Gerton, J. L., Ludtke, S. J., Akey, C. W., Chait, B. T., Sali, A., and Rout, M. P. (2018) Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475-482
59. Shaner, M. C., Blair, I. M., and Schneider, T. D. (1993) Sequence logos: A powerful, yet simple, tool. Proceedings of the twenty-sixth annual Hawaii international conference on system sciences. . in *Architecture and biotechnology computing* (al., T. N. M. e. ed.), IEEE Computer Society Press, Los Alamitos, CA. pp 813–821
60. Nei, M., and Li, W. H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A* **76**, 5269-5273

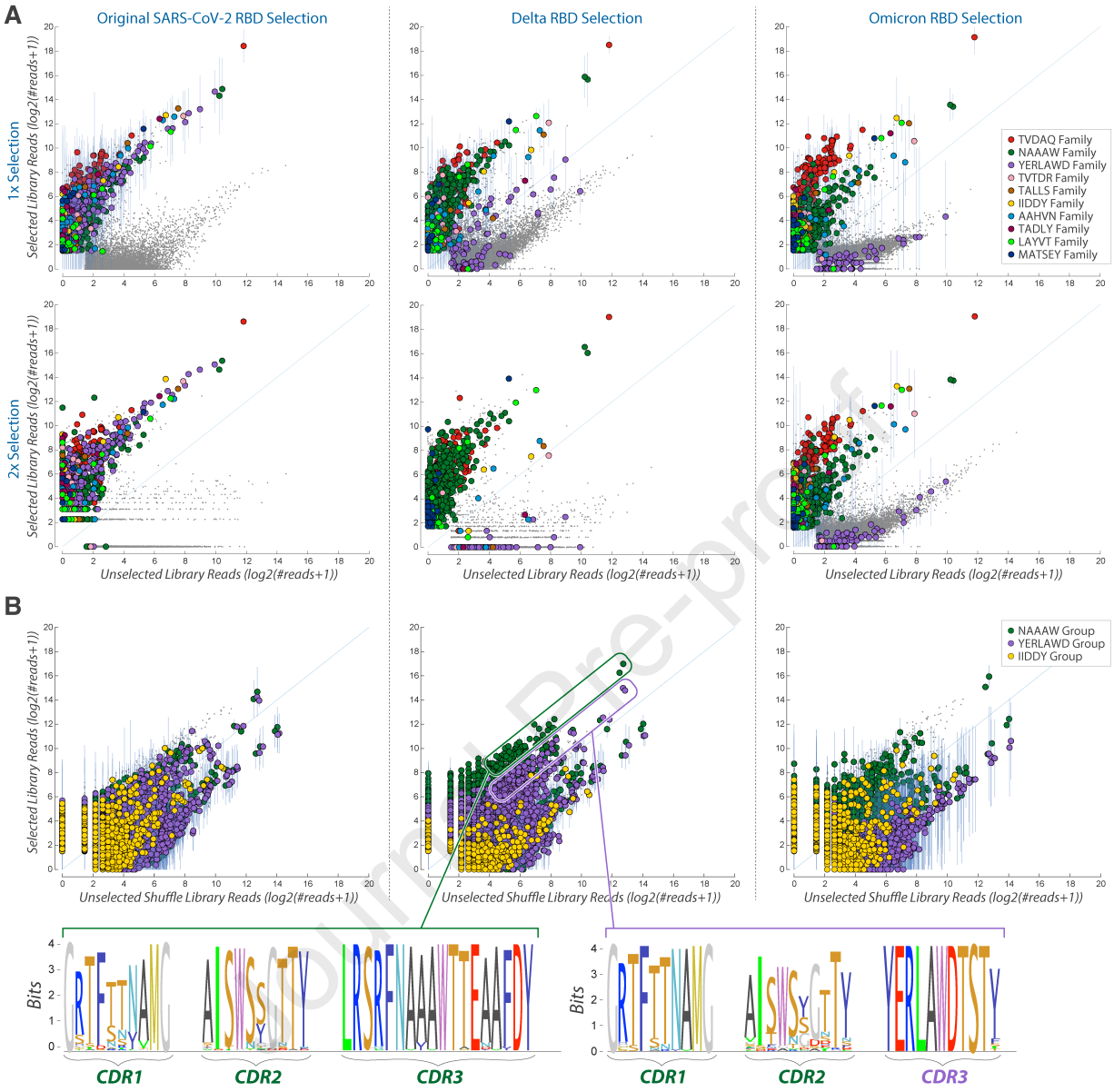


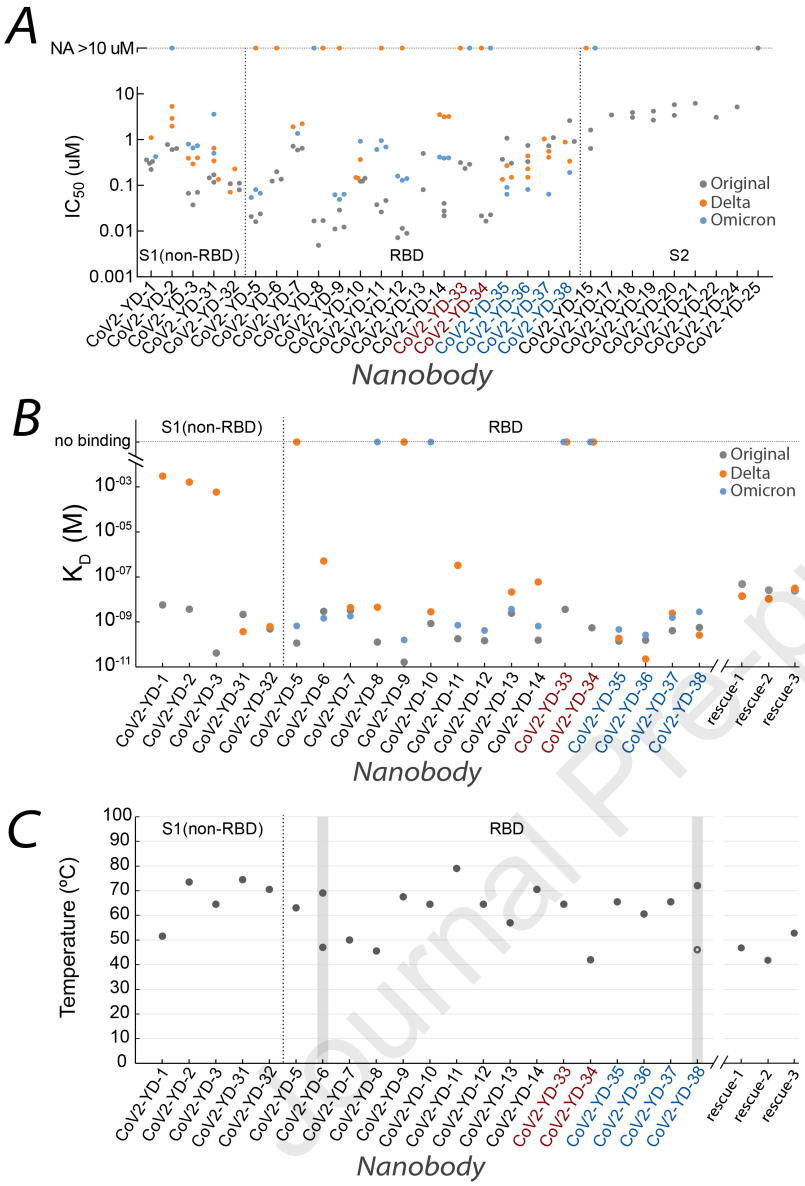


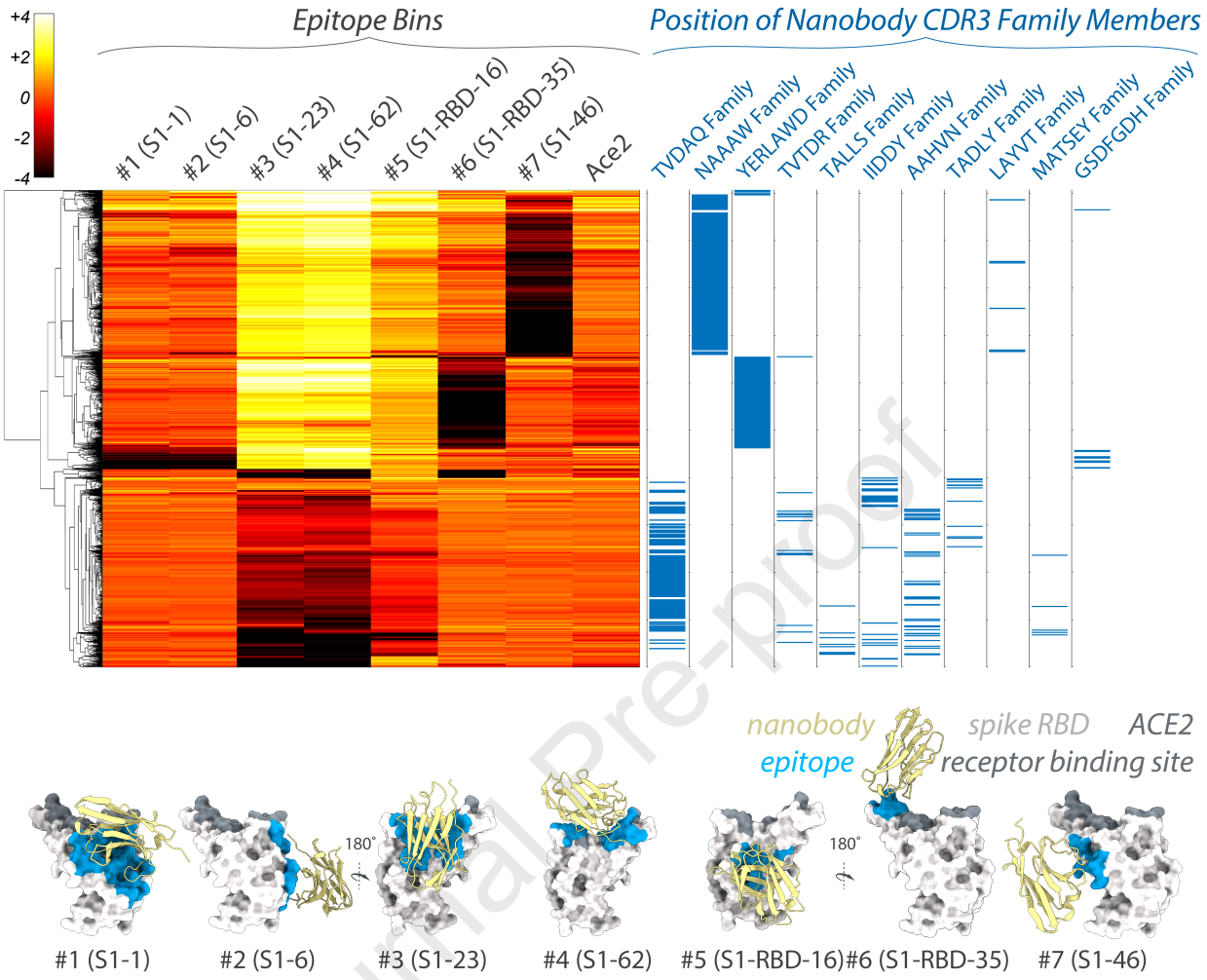












Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: