A library-on-library screen reveals the breadth expansion landscape of a broadly neutralizing betacoronavirus antibody

Marya Y. Ornelas[1], Wenhao O. Ouyang[2], Nicholas C. Wu[2,3,4,5,§]

[1] Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[2] Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[3] Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[4] Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[5] Carle Illinois College of Medicine, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[§] To whom correspondence may be addressed. Email: nicwu@illinois.edu (N.C.W.)

ABSTRACT

Broadly neutralizing antibodies (bnAbs) typically evolve cross-reactivity breadth through acquiring somatic hypermutations. While evolution of breadth requires improvement of binding to multiple antigenic variants, most experimental evolution platforms select against only one antigenic variant at a time. In this study, a yeast display library-on-library approach was applied to delineate the affinity maturation of a betacoronavirus bnAb, S2P6, against 27 spike stem helix peptides in a single experiment. Our results revealed that the binding affinity landscape of S2P6 varies among different stem helix peptides. However, somatic hypermutations that confer general improvement in binding affinity across different stem helix peptides could also be identified. We further showed that a key somatic hypermutation for breadth expansion involves long-range interaction. Overall, our work not only provides a proof-of-concept for using a library-on-library approach to analyze the evolution of antibody breadth, but also has important implications for the development of broadly protective vaccines.

INTRODUCTION

Many RNA viruses have high genetic diversity and undergo rapid antigenic drift due to the selection pressure from host humoral immune responses. As a result, it is often challenging to develop effective vaccines with a high breadth of protection against RNA viruses. Nevertheless, in the past two decades, a number of broadly neutralizing antibodies (bnAbs) against different RNA viruses, including influenza virus[1–4], human immunodeficiency virus[5–7], coronavirus[8–11], and flavivirus[12,13], have been isolated. These bnAbs protect against antigenically distinct strains of a given virus species, or even genus and family. The discovery of bnAbs has provided crucial insights into the development of broadly protective vaccines[9,14–16].

Most bnAbs evolve from an unmutated common ancestor (UCA) with narrow binding specificities[17–20]. Their cross-reactivity breadth is subsequently acquired through the accumulation of somatic hypermutations (SHMs) during affinity maturation[21–23]. Yeast display is a common approach for studying the evolutionary landscapes of antibody affinity maturation[24–32]. This process typically involves the construction of an antibody mutant library, which is then displayed on yeast surface and selected for antigen binding[24,26–28,30–32]. However, conventional yeast display selection focuses on only one antigen at a time, which imposes a challenge in studying the evolution of antibody breadth. Improving binding affinity of an antibody against one antigenic variant sometimes diminishes its binding affinity against another variant[27]. Yet, evolution of breadth requires improvement of binding affinity across multiple antigenic variants simultaneously. As a result, characterization of the evolutionary pathways for antibody breadth expansion requires selection against multiple antigenic variants in parallel.

In the past few years, studies of human antibody responses to SARS-CoV-2 have led to the discovery of bnAbs that target the highly conserved stem helix peptide in the S2 domain of the coronavirus spike glycoprotein[8,8,33,34]. S2P6, which was isolated from a COVID-19 convalescent

individual, is a representative bnAb to stem helix peptide[10]. While S2P6 neutralizes antigenically distinct betacoronavirus (β-CoV) strains including SARS-CoV-2, antibody binding data suggest that it arose in response to HCoV-OC43 infection and gained cross-reactivity to other β-CoV strains via SHMs[10]. Given that stem helix peptide is a target for the development of broadly protective coronavirus vaccines[35], it is important to understand the evolutionary trajectories that lead to breadth expansion of bnAbs to the stem helix peptide.

Recently, a yeast display platform for coevolving protein-protein interfaces was developed[36]. Here, we adopted this platform to screen a library of 27 unique β-CoV stem helix peptides against a mutant library of S2P6 encoding all combinations of SHMs that lie on or near the paratope. This approach enabled us to map the binding affinity landscapes of S2P6 against stem helix peptides from all β-CoV subgenera. We observed weak correlations of binding affinity landscapes of S2P6 across different β-CoV stem helix peptides, indicating that the effect of a given SHM on S2P6 binding could vary depending on the sequence of the target stem helix peptide. At the same time, several key SHMs for breadth expansion could be identified and experimentally validated. Our results further highlight the importance of long-range interaction in the affinity maturation of S2P6. Throughout this study, the Kabat numbering scheme is used for antibody residues unless otherwise stated.

RESULTS

Detecting interaction between S2P6 and SARS-CoV-2 stem helix peptide by yeast display

To determine if we could adopt a previously developed protein-protein coevolution platform[36] to screen an antibody mutant library against an antigen library, a pilot experiment was performed using the SARS-CoV-2 stem helix peptide and S2P6[10]. We cloned a yeast display construct that consisted of (from N-terminal to C-terminal): Aga2p, stem helix peptide (SP), 3C protease cleavage site and linker (3C), S2P6 in single-chain variable fragment (scFv) format, and HA tag

(**Fig. 1A and S1**). After the construct was displayed on the yeast surface, the yeast cells were treated with 3C protease. If S2P6 bound to the stem helix peptide, the HA tag would be retained at the yeast cell surface. If S2P6 did not bind to the stem helix peptide, the HA tag would be lost (**Fig. 1A**). For the yeast displaying the SARS-CoV-2 stem helix peptide and S2P6, the HA tag was readily detected both before and after protease treatment (**Fig. 1D**). This result indicated that S2P6 bound strongly to the SARS-CoV-2 stem helix peptide on yeast surface as expected. By contrast, for the yeast displaying the SARS-CoV-2 stem helix peptide and S2P6 UCA, which was the fully germline-reverted S2P6[10], we observed loss of the HA tag after protease treatment (**Fig. 1D**). This observation indicated that the S2P6 UCA did not bind to the SARS-CoV-2 stem helix peptide, consistent with the previous study[10]. Overall, our pilot experiment demonstrated that the interaction between S2P6 and stem helix peptide could be captured by the yeast display platform that was previously developed for studying protein-protein coevolution[36].

Screening an S2P6 mutant library against a stem helix peptide library

We sought to analyze the affinity maturation pathway of S2P6 by screening an S2P6 mutant library encoding all possible combinations of SHMs against a library of 27 unique β-CoV stem helix peptides (**Fig. 1B and Table S1**). Among the 12 SHMs in S2P6 (**Fig. 1C and S2A-B**), V_H Q1E and V_L D105E were distal from the binding interface. Reverting these positions in S2P6 to the germline sequence (V_H Q1 and V_L D105) did not result in loss of HA tag in our yeast display system after protease cleavage (**Fig. 1D**). This observation indicated that V_H Q1E and V_L D105E played a minimal, if any, role in the affinity maturation of S2P6. As a result, we moved forward with constructing the S2P6 mutant library without V_L D105E and V_H Q1E mutations (**Fig. S2C**). In other words, our S2P6 mutant library contained 1,024 variants (2^{10} = 1,024). Together with the 27 unique β-CoV stem helix peptides and one negative control peptide, there were a total of 28 peptide variants × 1,024 S2P6 variants = 28,672 combination variants in our library-on-library screen.

To track the frequency of different combination variants in our library, next-generation sequencing was required. While the region of interest in the previous study that developed the protein-protein coevolution platform was <600 bp[36], which is compatible with Illumina short-read sequencing, the amplicon spanning both S2P6 scFv and stem helix peptide was too long (>800 bp). To circumvent this issue, we adopted a barcoding strategy from a previous study[37]. Briefly, we included a 16-nucleotide "barcode" downstream of the coding region, such that different combination variants had their own barcodes (**Fig. S1 and S3**). The linkages between individual combination variants and their corresponding barcodes were then mapped by PacBio sequencing, which had sufficiently long read lengths to cover both S2P6 scFv and stem helix peptide but insufficient read depths to track the frequency of each combination variant in the library. Subsequently, the frequency of different variants could be tracked by sequencing the barcode region using Illumina short-read sequencing.

Flow cytometry analysis showed that combination variants in our library had a wide range of binding affinity (**Fig. 1D**). The library was subjected to fluorescence-activated cell sorting both before and after protease cleavage (**Fig. S4**). Four bins were collected from each sort according to the amount of HA tag detected on the yeast surface. The frequency of each combination variant in each bin was quantified by next-generation sequencing as described above. An expression score and a binding score were then computed for each combination variant based on its frequency distribution in different bins, hence a proxy for HA tag signal, before and after protease cleavage, respectively (**Table S2, see Materials and Methods**). As shown previously, binding affinity strongly correlates with HA tag signal after protease cleavage[36]. Combination variants in our library had a broad distribution of binding scores (**Fig. S5A**), consistent with the flow cytometry analysis of the library (**Fig. 1D**). At the same time, they had a narrow distribution of expression scores (**Fig. S5B**), indicating that most combination variants expressed well. We also observed a

Pearson correlation of 0.70 between the binding scores of two independent experimental replicates (**Fig. S5C**), demonstrating reproducibility of our library-on-library screen. By contrast, the Pearson correlation between binding score and expression score was 0.29, showing that the expression levels of combination variants had a relatively small influence on their binding scores (**Fig. S5D**).

Breadth expansion of S2P6 is restricted by sequence variations of stem helix peptides

To analyze the breadth expansion of S2P6, we computed the average binding scores of the S2P6 mutant library for each β-CoV stem helix peptide (**Fig. 2A**). High average binding scores were observed for stem helix peptides from all sarbecovirus strains as well as embecovirus strains, except PHEV and HCoV-HKU1. Nevertheless, the average binding scores for stem helix peptides from PHEV and HCoV-HKU1 were still higher than the negative control peptide. For stem helix peptides from the embecovirus, amino acid variants at position 1153 appeared to play a critical role in binding to S2P6. At position 1153, YakCoV, HCoV-HKU1, and PHEV have N1153, S1153, and Y1153, respectively, whereas all other embecovirus strains have D1153 **(Fig. 1B)**. As previously shown by a mutational analysis of SARS-CoV-2 stem helix peptide, D1153 has the highest binding activity to S2P6, followed by N1153, S1153, and then Y1153[10]. Consistently, stem helix peptides from those embecovirus strains with D1153 had the highest average binding scores in our data, followed by YakCoV (N1153), HCoV-HKU1 (S1153), then PHEV (Y1153).

Low average binding scores were observed for stem helix peptides from all merbecovirus strains, except MERS-CoV and Bat-NeoCoV. Notably, both MERS-CoV and Bat-NeoCoV have D1153 whereas all other merbecovirus strains have E1153 (**Fig. 1B**), again substantiating the importance of Asp at this position for S2P6 binding[10]. All hibecovirus strains also have E1153, which explained the low average binding scores to their stem helix peptides. Furthermore, stem helix peptides from all nobecovirus strains had low average binding scores. While the stem helix peptides from

other β-CoV strains have L1152, all nobecovirus strains have F1152 (**Fig. 1B**), which has previously been shown to abolish binding to S2P6[10].

S2P6 binding landscape varies across stem helix peptides

It is known that S2P6 UCA binds strongly to the stem helix peptides of HCoV-OC43, but not those from the other two β-CoV strains that circulate in human population, namely HCoV-HKU1 and SARS-CoV-2[10]. As a result, S2P6 is hypothesized to have been initially elicited in response to HCoV-OC43 infection with specificity broadened through subsequent recall response to either or both HCoV-HKU1 and SARS-CoV-2 infections[10]. Consistently, even S2P6 variants with a low number of SHMs had high binding scores for HCoV-OC43 stem helix peptide (**Fig. 2B**). By contrast, the binding scores for stem helix peptides from all sarbecovirus strains, namely SARS-CoV-2, BatCoV-BtkY72, BtRf-BetaCoV, and BatCoV-GX2013, increased as S2P6 accumulated more SHMs (**Fig. 2B and S6**).

We further aimed to compare the binding landscapes of S2P6 across different β-CoV stem helix peptides (**Fig. 3 and S7**). Weak correlations of binding scores were observed among stem helix peptides from sarbecovirus strains (Pearson correlation = 0.14 to 0.21, **Fig. 3**). There was also a weak correlation of binding scores between BCoV-ENT (embecovirus) and sarbecovirus stem helix peptides (Pearson correlation = 0.12 to 0.26). In comparison, the correlations of binding scores between experimental replicates for each of these stem helix peptides were higher (Pearson correlation = 0.35 to 0.41, **Fig. 3**). Similarly, there was a lack of correlation of binding scores among embecovirus and merbecovirus stem helix peptides (Pearson correlation = -0.13 to 0.08), whereas moderate correlations of binding scores were observed between their experimental replicates (Pearson correlation = 0.34 to 0.56). These results demonstrated that the S2P6 binding landscape differed among stem helix peptides from different β-CoV strains. In other

187 words, depending on the sequence of the target stem helix peptide, a given SHM of S2P6 could
188 have different effects on binding affinity.
189
190 **Affinity maturation of S2P6 involves long range interaction**
191 Next, we aimed to analyze the impacts of individual SHMs on S2P6 binding landscape. Briefly,
192 the S2P6 binding landscape was decomposed into additive effects of individual SHMs and
193 pairwise epistasis effects between SHMs (**Fig. 4A and S8, see Materials and Methods**)[38]. Three
194 SHMs on heavy chain, namely V_H Y32Q, V_H M48I, and V_H S56H, stood out as having positive
195 additive effects on binding to stem helix peptides from most β-CoV strains (**Fig. 4A**). To
196 experimentally validate this finding, these mutations were introduced individually and in
197 combinations (single, double, and triple mutants) to S2P6 UCA and expressed as fragment
198 antigen-bindings (Fabs). Biolayer interferometry experiments showed that each single mutation
199 improved the binding response to the stem helix peptides from BatCoV-BtkY72, SARS-CoV-2,
200 BtRf-BetaCoV, and MHV in comparison to S2P6 UCA (**Fig. 4B and S9**). Consistently, reverting
201 V_H Q32 to V_H Y32 has been shown to decrease the binding affinity of S2P6 to multiple β-CoV
202 stem helix peptides[10]. Nevertheless, among the three SHMs tested, V_H M48I conferred the largest
203 improvement in binding response to all four tested stem helix peptides. V_H M48I alone led to a
204 higher binding response than V_H Y32Q/S56H and V_H Y32Q/M48I double mutants for BatCoV-
205 BtkY72, SARS-CoV-2, and BtRf-BetaCoV stem helix peptides. Similarly, for the stem helix
206 peptides from BatCoV-BtkY72, SARS-CoV-2, and BtRf-BetaCoV, V_H M48I/S56H double mutant
207 had higher binding response than the V_H Y32Q/M48I/S56H triple mutant. Since V_H Y32Q did not
208 improve the binding response in the presence of V_H M48I, our data also suggested that negative
209 epistasis existed between V_H Y32Q and V_H M48I, which was also observed in our decomposition
210 analysis, albeit mildly (**Fig. S8**).
211

Since V_H Y32Q and V_H S56H are in the paratope of S2P6[10] (**Fig. 4C**), it may not be surprising that they could improve the binding of S2P6 to β-CoV stem helix peptides. By contrast, V_H M48I is not in the paratope and instead resides in the hydrophobic core of the heavy chain variable domain (**Fig. 4C**). This observation indicated that V_H M48I improved the interaction between S2P6 and β-CoV stem helix peptides via long-range interaction. In fact, it is not uncommon for non-paratope SHMs to improve antibody binding affinity by modulating the conformation of complementarity determining region (CDR) loops that are part of the paratope[17,39]. It is possible that a similar mechanism applies to V_H M48I in S2P6, although additional structural analysis is required to verify this speculation.

DISCUSSION

Conventional yeast display experiments enable high-throughput analysis of antibody affinity maturation pathways[28,30,31,40]. However, they are often limited to analyzing antibody binding to one target antigen at a time, which restricts our understanding in the evolution of cross-reactivity breadth. A platform that allows analysis of affinity maturation pathways to multiple antigenic variants in parallel is therefore needed. In this study, we addressed this gap by determining the binding affinity landscapes of S2P6[10] against 27 unique stem helix peptides across all β-CoV subgenera in a single yeast display experiment. Our work provides not only a proof-of-concept for using a yeast display library-on-library approach to analyze the evolution of antibody cross-reactivity breadth, but also molecular insights into the development of broadly protective coronavirus vaccines.

A notable finding in this study is the weak correlations of S2P6 binding landscapes among different β-CoV stem helix peptides. This indicates the varying effects of a given SHM on binding affinity to different β-CoV stem helix peptides. Consistently, our decomposition analysis showed that most SHMs had varying additive effects. Similar observations have also been made with an influenza

bnAb in a previous study[30]. Nevertheless, we also identified SHMs that improve the binding of S2P6 to most β-CoV stem helix peptides. An effective broadly protective vaccine should facilitate the acquisition of SHMs that improve binding affinity to many antigenic variants. On the other hand, SHMs that confer trade-offs for binding different antigenic variants should be avoided. Consequently, future studies need to focus on the structural mechanisms underlying the differential impacts of SHMs on breadth expansion, which in turn will be informative for the development of vaccines that optimally elicit bnAbs.

Using library-on-library approaches to screen antibody-antigen interactions has been described since at least 15 years ago, based on co-selection of a phage-displayed antibody library and a yeast-displayed antigen library[41]. Since this strategy requires cell sorting into 96-well plates followed by sequencing the antibody and antigen in each well individually[41], its throughput is low. More recently, library-on-library approaches that utilize lentivirus-displayed antigen library and human B cell library have been described[42,43]. However, the throughput of these approaches remains moderate due to their reliance on single-cell sequencing, which analyzes at most ~10^4 cells per sample. By contrast, the yeast display library-on-library approach in our present study, which was adopted from Yang et al.[36], has a much higher throughput since it is compatible with amplicon-based next-generation sequencing. If nucleotide barcode is needed as described in this study, the maximum throughput would be ~10^6, as determined by the throughput of a PacBio sequencing run. In fact, another yeast display library-on-library approach has been described previously that relies on yeast mating and is also compatible with amplicon-based next-generation sequencing[44]. Nevertheless, this approach requires an engineered yeast strain that is not commercially available[44]. In comparison, our approach is more accessible since it largely follows conventional yeast display protocol without the need of any special reagents.

We acknowledge that our proof-of-concept in this study is based on a small peptide antigen. However, given that larger antigens, such as influenza hemagglutinin, HIV gp120, and SARS-CoV-2 receptor-binding domain, have been successfully displayed on the yeast surface[45–47], we anticipate that our library-on-library approach can be applied to a wide range of antibody-antigen pairs. In addition, since our approach does not rely on yeast biology, unlike the yeast mating-based approach mentioned above[44], it can potentially be adopted to a mammalian cell display system, which is essential for even larger antigens or if glycosylation matters. Future studies should continue to improve existing library-on-library approaches for screening antibody-antigen interactions as well as develop new ones.

MATERIALS AND METHODS

Yeast display plasmid

A DNA fragment encoding (from N-terminal to C-terminal) a cMyc tag, SARS-CoV-2 spike stem helix peptide sequence (residues 1146-1159), 3C protease cleavage site and linker, S2P6 heavy chain variable domain (V_H) and light chain variable domain (V_L) connected with a GS flexible linker, and an HA tag was synthesized as an eBlock (Integrated DNA Technologies). Subsequently, this fragment was cloned into the pCTcon2 vector (Addgene, Cat. No. 41843)[48] in frame with the Aga2 coding region at the N-terminal end using Gibson assembly (**Fig. S1**).

Construction of the combination variant library

The cMyc-SP-3C-S2P6_VH-GS-S2P6_VL-HA library insert was divided into six fragments and assembled using overlap PCR (**Fig. S3**). All fragments were gel-purified using a Monarch DNA Gel Extraction Kit (NEB). To generate the full-length library insert, all fragments were pooled together and underwent 10 cycles of PCR with the following conditions: 98 °C for 30 s, 10 cycles of (98 °C for 10 s, 50 °C for 15 s, 72 °C for 15 s), 72 °C for 2 min, 4 °C indefinitely. Extension primers Fragment 1_F and Fragment 6_R_3 were added into the reaction, which were then ran

for another 25 cycles using conditions: 98 °C for 30 s, 25 cycles of (98 °C for 10 s, 55 °C for 5 s, 72 °C for 15 s), 72 °C for 2 min, 4 °C indefinitely. The full-length library insert was gel-purified. To generate the linearized vector for the library, the plasmid encoding the SARS-CoV-2 stem helix peptide and S2P6 WT, which was described above, was used as a template for PCR using primers S2P6_vector_F and S2P6_vector_R. The linearized vector was digested with DpnI (NEB) for 2 h at 37 °C and gel purified. All PCRs were performed using PrimeSTAR Max polymerase (Takara Bio). The sequences of all primers in this study can be found in **Table S3**. All primers in this study were synthesized by Integrated DNA Technologies.

Yeast transformation

Saccharomyces cerevisiae EBY100 cells (American Type Culture Collection, Cat. No. MYA-4941) were grown overnight in YPD medium (1% w/v yeast nitrogen base, 2% w/v peptone, 2% w/v D(+)-glucose) at 30 °C with shaking at 225 rpm until OD_{600} reached 3. An aliquot of the overnight culture was used to inoculate 100 mL fresh YPD media with an initial OD_{600} of 0.3. This culture was incubated at 30 °C with shaking at 225 rpm until OD_{600} reached 1.6. The yeast cells were collected by centrifugation at 4600 × g for 3 min at room temperature. YPD media was removed, and the cell pellet was washed twice with 50 mL ice-cold water, and then once with 50 mL of ice-cold electroporation buffer (1 M sorbitol and 1 mM calcium chloride). The cells were then resuspended in 20 mL conditioning media (0.1 M lithium acetate and 10 mM dithiothreitol) and incubated at 30 °C with shaking at 225 rpm. The conditioned yeast cells were then collected via centrifugation at 4600 × g for 3 min at room temperature, washed once with 50 mL ice-cold electroporation buffer and resuspended in electroporation buffer to reach a final volume of 1 mL. The cells were kept on ice until used.

5 µg of the purified linearized vector and 2.7 µg of the purified library insert were added to 400 µL of conditioned yeast. The mixture was transferred to a pre-chilled BioRad GenePulser cuvette

with 2 mm electrode gap and kept on ice for 5 min prior to electroporation. Cells were electroporated at 2.5 kV and 25 μF, achieving a time constant between 3.7 and 4.1 ms. Electroporated cells were recovered into 4 mL of YPD media supplemented with 4 mL of 1 M sorbitol and incubated at 30 °C with shaking at 225 rpm for 1 h. Recovered cells were collected via centrifugation at 1700 × g for 3 min at room temperature, resuspended in 0.6 mL SD-CAA medium (2% w/v D-glucose, 0.67% w/v yeast nitrogen base with ammonium sulfate, 0.5% w/v casamino acids, 0.54% w/v Na_2HPO_4, and 0.86% w/v $NaH_2PO_4 \cdot H_2O$, all dissolved in deionized water), plated across 3× 150 mm SD-CAA plates (2% w/v D-glucose, 0.67% w/v yeast nitrogen base with ammonium sulfate, 0.5% w/v casamino acids, 0.54% w/v Na_2HPO_4, 0.86% w/v $NaH_2PO_4 \cdot H_2O$, 18.2% w/v sorbitol, and 1.5% w/v agar, all dissolved in deionized water) and incubated at 30 °C for 48 h. After 48 h, approximately 200,000 colonies were collected in SD-CAA medium, centrifuged at 4600 × g for 5 min at room temperature, and resuspended in SD-CAA medium with 15% v/v glycerol such that OD_{600} was 50. Glycerol stocks were stored at −80 °C for later use. This protocol was modified from a previously described protocol[49].

PacBio sequencing of the combination variant library

Plasmids from the transformed yeast cells were extracted using a Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research) following the manufacturer's protocol. The library insert and barcode were amplified using primers PacBio Recovery_F and PacBio Recovery_R (**Table S3**). This PCR was performed using Q5 high-fidelity DNA polymerase (NEB) with the following settings: 98 °C for 30 s, 25 cycles of (98 °C for 10 s, 62 °C for 15 s, 72 °C for 30 s), 72 °C for 2 min, 4 °C indefinitely. PCR products were gel purified and sequenced on one SMRT Cell 8M on a PacBio Sequel IIe using the CCS sequencing mode and a 15-hour movie time.

Analysis of PacBio sequencing data

Circular consensus sequences (CCSs) were generated from the raw subreads using the ccs program (https://github.com/PacificBiosciences/ccs, version 6.4.0), setting the parameters to require 99.9% accuracy and a minimum of 3 passes. CSSs in FASTQ format were parsed using SeqIO module in BioPython[50]. For each read, the positions of the stem helix peptide, S2P6 heavy and light chain variable domains, as well as the 16-nucleotide barcode were identified by aligning their flanking sequences to the read. If the lengths of these regions deviated from the expected lengths, the read would be discarded. The nucleotide sequences of the stem helix peptide, as well as S2P6 heavy and light chain variable domains were then translated. If the amino acid sequence of the stem helix peptide did not match any of the 27 stem helix peptides, the read would be discarded. Similarly, if the amino acid sequences at the mutated positions of S2P6 did not match our library design, the read would be discarded. Subsequently, the combination variant and the 16-nucleotide barcode in each read were identified.

Since reads that shared the same 16-nucleotide barcode should have the same combination variant, we compared the amino acid sequences at the positions of interest among different reads that shared the same 16-nucleotide barcode. For each position of interest, an amino acid variant present in at least 80% of the reads that shared the same 16-nucleotide barcode was regarded as the true variant. If none of the amino acid variant was present in at least 80% of the reads that shared the same 16-nucleotide barcode, all reads that shared the given 16-nucleotide barcode would be discarded. Besides, those 16-nucleotide barcodes that appeared in only one read were discarded.

Fluorescence-activated cell sorting (FACS) of yeast display library

A 150 µL glycerol stock of the transformed yeast display library was recovered in 50 mL SD-CAA medium by incubating at 27 °C with shaking at 250 rpm until OD_{600} reached between 1.5-2.0. Then 15 mL of the yeast culture was harvested via centrifugation at 4600 × g at 4 °C for 5 min.

366 The supernatant was discarded, and SGR-CAA induction media (2% w/v galactose, 2% w/v
367 raffinose, 0.1% w/v D-glucose, 0.67% w/v yeast nitrogen base with ammonium sulfate, 0.5% w/v
368 casamino acids, 0.54% w/v Na_2HPO_4, and 0.86% w/v $NaH_2PO_4 \cdot H_2O$, all dissolved in deionized
369 water) was added to a final volume of 50 mL with an initial OD_{600} of 0.5. The SGR-CAA yeast
370 culture was transferred to a baffled flask and incubated at 18 °C with shaking at 250 rpm for about
371 24 h until the OD_{600} reached between 1.3-1.6.
372
373 For each staining replicate, 1 mL of yeast culture was harvested via centrifugation at 4600 × g at
374 4 °C for 5 min. The pellet was washed twice with 1 mL of 1× PBS and resuspended in 1 mL of 1×
375 PBS. For the library binding sort, anti-HA-tag mouse antibody with Alexa Fluor 647 conjugate
376 (6E2, Cell Signaling Technology, Cat. No. 3444) was added to washed cells at a 1:100 dilution. A
377 no-stain negative control was included in which nothing was added to the PBS-resuspended cells.
378 Samples were incubated overnight at 4 °C with rotation. For protease treatment, the stained cells
379 were spun down and the supernatant was removed. 10 µL HRV 3C protease (ThermoFischer,
380 Cat. No. 88946), 10 µL protease reaction buffer, and 80 µL water was added to the stained cells.
381 The cells incubated with protease at 4 °C for 1 h.
382
383 The yeast cells were pelleted and washed twice in 1× PBS and resuspended in FACS tubes
384 containing 2 mL 1× PBS. Using a BD FACS Aria II cell sorter (BD Biosciences) and FACS Diva
385 software v8.0.1 (BD Biosciences), cells in the selected gates were collected in 1 mL of SD-CAA
386 containing 1× penicillin/streptomycin. Single yeast cells were gated by forward scatter (FSC) and
387 side scatter (SSC). For the binding sort, single cells were gated into 4 bins along the Alexa Fluor
388 647 axis. Cells expressing the highest Alexa Fluor 647 fluorescence were sorted into “bin 4”, then
389 the next highest into “bin 3”, followed by “bin 2” and then “bin 1”. Gating strategy used is shown
390 in **Fig. S4A**. Number of cells collected per bin per replicate is shown in **Fig. S4B**. Cells were then
391 collected and grown overnight in SD-CAA at 30 °C with shaking at 225 rpm. FlowJo v10.8 software

(BD Life Sciences) was used to analyze FACS data. Replicates were performed on the same day, starting from separate glycerol stocks of the transformed library.

Illumina sequencing of the post-sorted yeast display library

Plasmids from the sorted yeast cells were extracted using a Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research) following the manufacturer's protocol. The 16-nucleotide barcode was amplified using primers Barcode_Recovery_F and Barcode_Recovery_R (**Table S3**), which contained part of the adapter sequence required for Illumina sequencing. A maximum of 100 ng of genomic DNA per 50 µL PCR reaction was used as template. This PCR was performed using Q5 high-fidelity DNA polymerase (NEB) with the following settings: 98 °C for 30 s, 25 cycles of (98 °C for 10 s, 62 °C for 15 s, 72 °C for 15 s), 72 °C for 2 min, 4 °C indefinitely. The PCR products were gel purified. For each sample, 20 ng of the purified PCR product was appended with the rest of adapter sequence and index via PCR using primers: 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACX XXX XXX XAC ACT CTT TCC CTA CAC GAC GCT-3', and 5'-CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX XXG TGA CTG GAG TTC AGA CGT GTG CT-3'. Positions annotated by an "X" represented the nucleotides for the index sequence. This PCR was performed by PrimeSTAR Max polymerase (Takara Bio) with the following settings: 98 °C for 30 s, 10 cycles of (98 °C for 10 s, 55 °C for 15 s, 72 °C for 15 s), 72 °C for 2 min, 4 °C indefinitely. Indexed products were gel purified, mixed at equimolar ratios, and submitted for next generation sequencing using Illumina NovaSeq X PE150.

Analysis of Illumina sequencing data

The Illumina NovaSeq sequencing data were obtained in FASTQ format. Forward and reverse reads of each paired-end read were merged by PEAR[51]. The merged reads were then parsed by SeqIO module in BioPython[50], followed by primer trimming to remove primer sequences. Reads with incorrect primer sequences or lengths were discarded. Reads that did not match any

barcodes from our PacBio data were also discarded. The combination variant, which consisted of an S2P6 variant *i* and a stem helix peptide variant *s,* in each read was then identified using the barcode. The frequency (*F*) of a combination variant within bin *n* at timepoint *t* of replicate *r* was computed as follows:

$$F_{s,i,n,t,r} = \frac{readcount_{s,i,n,t,r}+1}{\sum_i \sum_s (readcount_{s,i,n,t,r}+1)} \quad (1)$$

A pseudocount of 1 was added to the read counts of each mutant to avoid division by zero in subsequent steps. The frequency (*F*) of each combination variant among bins at *t* = 0 h (i.e. pre-protease cleavage) was used to calculate the expression score (*E*)[37,52]:

$$E_{s,i,r} = \frac{F_{s,i,bin1,t_0,r} \times 0.25 + F_{s,i,bin2,t_0,r} \times 0.5 + F_{s,i,bin3,t_0,r} \times 0.75 + F_{s,i,bin4,t_0,r} \times 1}{\sum_{n=1}^{4} F_{s,i,n,t_0,r}} \quad (2)$$

The frequency (*F*) of each combination variant among bins at *t* = 1 h (i.e. post-protease cleavage) was then used to calculate the binding score (*B*):

$$B_{s,i,r} = \frac{F_{s,i,bin1,t_1,r} \times 0.25 + F_{s,i,bin2,t_1,r} \times 0.5 + F_{s,i,bin3,t_1,r} \times 0.75 + F_{s,i,bin4,t_1,r} \times 1}{\sum_{n=1}^{4} F_{s,i,n,t_1,r}} \quad (3)$$

The average binding score for combination variants that contained the negative control peptide (DSAKEALDKYFKNH) were calculated ($B_{negctrl}$). The binding score of the combination variant that contained SARS-CoV-2 stem helix peptide (DSFKEELDKYFKNH) and S2P6 WT was also extracted ($B_{SARS2\text{-}S2P6WT}$). The normalized binding score (*NormB*) of each combination variant *i* was then calculated as followed:

$$NormB_{s,i,r} = \frac{B_{s,i,r} - \overline{B_{negctrl,r}}}{B_{SARS2-S2P6WT,,r} - \overline{B_{negctrl,r}}} \quad (4)$$

The reported *NormB* was the average between replicates.

Expression and purification of Fabs

Codon-optimized oligonucleotides encoding S2P6 V_H variants were synthesized as eBlocks (Integrated DNA Technologies), and then cloned into phCMV3 vector in Fab format using Gibson assembly. The plasmids were co-transfected into Expi293F cells at a 2:1 (HC:LC) mass ratio using ExpiFectamine 293 Reagent (Thermo Fisher Scientific) following the manufacturer's protocol for a 25 mL culture. 6 days post-transfection, the supernatant of each Expi293F culture was collected and clarified by centrifugation at 4600 × g for 45 min at 4 °C to remove cells and debris. After clarification, the supernatant was filtered through a 0.22 µm polyethersulfone membrane filter (Millipore).

Fabs were purified using IgG-CH1-XL beads (Thermo Fisher Scientific). The beads were washed with MilliQ H_2O three times and resuspended in 1× PBS. The clarified and filtered supernatant was incubated with washed beads overnight at 4 °C with gentle rocking. The supernatant with beads was then loaded into a column. The flowthrough was discarded, and the beads were washed once with 1× PBS. To elute the antibody, the beads were incubated in 60 mM sodium acetate, pH 3.7 for 10 min at 4 °C and the flow-through was collected. The eluate containing Fab was buffer exchanged into 1× PBS using a centrifugal filter unit with a 10 kDa molecular weight cut-off (Millipore). Fabs were stored at 4 °C.

Biolayer interferometry binding assays

Biolayer interferometry (BLI) experiments were performed at room temperature using Octet Red96e instrument (Sartorius). Biotin-labeled stem peptides (GenScript) at 4 µg/mL in 1× kinetics buffer (1× PBS, pH 7.4, 0.01% w/v BSA, and 0.002% v/v Tween 20) were loaded onto streptavidin (SA) biosensors and incubated with Fabs at 500 nM, 750 nM, and 1000 nM. The assay consisted of five steps: (1) baseline: 60 s with 1× kinetics buffer; (2) loading: 120 s with biotin-labeled stem peptides; (3) baseline: 60 s with 1× kinetics buffer; (4) association: 120 s with Fab samples; and (5) dissociation: 120 s with 1× kinetics buffer. For estimating the exact K_D, a 1:1 binding model

was used. When a 1:1 binding model did not fit well due to non-specific binding, a 2:1 heterogeneous ligand model was used to improve fitting.

DATA AVAILABILITY

Raw deep sequencing data generated in this study have been submitted to the NIH Sequence Read Archive under accession number: PRJNA1064076 and PRJNA1113356.

CODE AVAILABILITY

Custom codes for data analysis have been deposited to: https://github.com/nicwulab/S2P6_lib-on-lib_screen.

ACKNOWLEDGMENTS

We thank the Roy J. Carver Biotechnology Center at the University of Illinois Urbana-Champaign for assistance with fluorescence-activated cell sorting and performing next-generation sequencing. This work was supported by the US National Institutes of Health DP2 AT011966 (N.C.W.), R01 AI167910 (N.C.W.), Searle Scholars Program (N.C.W.), National Institutes of the Health Chemical Biology Interface Training Program T32 GM136629 (M.Y.O.), and National Science Foundation Graduate Research Fellowship DGE 21 46756 (M.Y.O.).

AUTHOR CONTRIBUTIONS

M.Y.O. and N.C.W. designed research. M.Y.O. performed research. M.Y.O., W.O.O., and N.C.W. analyzed data. M.Y.O. and N.C.W. wrote the paper.

COMPETING INTERESTS

N.C.W. serves as a consultant for HeliXon. All authors declare no other competing interests.

REFERENCES

1. Dreyfus, C. *et al.* Highly conserved protective epitopes on influenza B viruses. *Science* **337**, 1343–1348 (2012).
2. Guthmiller, J. J. *et al.* Broadly neutralizing antibodies target a haemagglutinin anchor epitope. *Nature* **602**, 314–320 (2022).
3. Sun, X. *et al.* Broadly neutralizing antibodies to combat influenza virus infection. *Antiviral Res.* **221**, 105785 (2024).
4. Laursen, N. S. & Wilson, I. A. Broadly neutralizing antibodies against influenza viruses. *Antiviral Res.* **98**, 476–483 (2013).
5. Hodge, E. A. *et al.* An HIV-1 broadly neutralizing antibody overcomes structural and dynamic variation through highly focused epitope targeting. *Npj Viruses* **1**, 1–11 (2023).
6. Liu, Y., Cao, W., Sun, M. & Li, T. Broadly neutralizing antibodies for HIV-1: efficacies, challenges and opportunities. *Emerg. Microbes Infect.* **9**, 194–206.
7. Paneerselvam, N., Khan, A. & Lawson, B. R. Broadly neutralizing antibodies targeting HIV: progress and challenges. *Clin. Immunol.* **257**, 109809 (2023).
8. Zhou, P. *et al.* Broadly neutralizing anti-S2 antibodies protect against all three human betacoronaviruses that cause deadly disease. *Immunity* **56**, 669-686.e7 (2023).
9. Chen, Y. *et al.* Broadly neutralizing antibodies to SARS-CoV-2 and other human coronaviruses. *Nat. Rev. Immunol.* **23**, 189–199 (2023).
10. Pinto, D. *et al.* Broad betacoronavirus neutralization by a stem helix–specific human antibody. *Science* **373**, 1109–1116 (2021).
11. He, W. *et al.* Targeted isolation of diverse human protective broadly neutralizing antibodies against SARS-like viruses. *Nat. Immunol.* **23**, 960–970 (2022).
12. Deng, Y.-Q. *et al.* A broadly flavivirus cross-neutralizing monoclonal antibody that recognizes a novel epitope within the fusion loop of E protein. *PLoS ONE* **6**, e16059 (2011).

13. Pierson, T. C. & Diamond, M. S. The continued threat of emerging flaviviruses. *Nat. Microbiol.* **5**, 796–812 (2020).

14. Sok, D. & Burton, D. R. Recent progress in broadly neutralizing antibodies to HIV. *Nat. Immunol.* **19**, 1179–1188 (2018).

15. Haynes, B. F. *et al.* Strategies for HIV-1 vaccines that induce broadly neutralizing antibodies. *Nat. Rev. Immunol.* **23**, 142–158 (2023).

16. Wu, N. C. & Wilson, I. A. Structural insights into the design of novel anti-influenza therapies. *Nat. Struct. Mol. Biol.* **25**, 115–121 (2018).

17. Wu, N. C. *et al.* Convergent evolution in breadth of two VH6-1-encoded influenza antibody clonotypes from a single donor. *Cell Host Microbe* **28**, 434-444.e4 (2020).

18. Bhiman, J. N. *et al.* Viral variants that initiate and drive maturation of V1V2-directed HIV-1 broadly neutralizing antibodies. *Nat. Med.* **21**, 1332–1336 (2015).

19. Griffith, S. A. & McCoy, L. E. To bnAb or not to bnAb: defining broadly neutralising antibodies against HIV-1. *Front. Immunol.* **12**, 708227 (2021).

20. Derdeyn, C. A., Moore, P. L. & Morris, L. Development of broadly neutralizing antibodies from autologous neutralizing antibody responses. *Curr. Opin. HIV AIDS* **9**, 210–216 (2014).

21. Mishra, A. K. & Mariuzza, R. A. Insights into the structural basis of antibody affinity maturation from next-generation sequencing. *Front. Immunol.* **9**, 117 (2018).

22. Rajewsky, K. Clonal selection and learning in the antibody system. *Nature* **381**, 751–758 (1996).

23. Teng, G. & Papavasiliou, F. N. Immunoglobulin somatic hypermutation. *Annu. Rev. Genet.* **41**, 107–120 (2007).

24. Koenig, P. *et al.* Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding. *Proc. Natl. Acad. Sci.* **114**, E486–E495 (2017).

25. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).

26. Adams, R. M., Mora, T., Walczak, A. M. & Kinney, J. B. Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *eLife* **5**, e23156 (2016).

27. Wu, N. C. *et al.* In vitro evolution of an influenza broadly neutralizing antibody is modulated by hemagglutinin receptor specificity. *Nat. Commun.* **8**, 15371 (2017).

28. Madan, B. *et al.* Mutational fitness landscapes reveal genetic and structural improvement pathways for a vaccine-elicited HIV-1 broadly neutralizing antibody. *Proc. Natl. Acad. Sci.* **118**, e2011653118 (2021).

29. Wellner, A. *et al.* Rapid generation of potent antibodies by autonomous hypermutation in yeast. *Nat. Chem. Biol.* **17**, 1057–1064 (2021).

30. Phillips, A. M. *et al.* Binding affinity landscapes constrain the evolution of broadly neutralizing anti-influenza antibodies. *eLife* **10**, e71393 (2021).

31. Phillips, A. M. *et al.* Hierarchical sequence-affinity landscapes shape the evolution of breadth in an anti-influenza receptor binding site antibody. *eLife* **12**, e83628 (2023).

32. Banach, B. B. *et al.* Antibody-directed evolution reveals a mechanism for enhanced neutralization at the HIV-1 fusion peptide site. *Nat. Commun.* **14**, 7593 (2023).

33. Zhou, P. *et al.* A human antibody reveals a conserved site on beta-coronavirus spike proteins and confers protection against SARS-CoV-2 infection. *Sci. Transl. Med.* **14**, eabi9215 (2022).

34. Dacon, C. *et al.* Rare, convergent antibodies targeting the stem helix broadly neutralize diverse betacoronaviruses. *Cell Host Microbe* **31**, 97-111.e12 (2023).

35. Kapingidza, A. B. *et al.* Engineered immunogens to elicit antibodies against conserved coronavirus epitopes. *Nat. Commun.* **14**, 7897 (2023).

36. Yang, A. *et al.* Deploying synthetic coevolution and machine learning to engineer protein-protein interactions. *Science* **381**, eadh1720 (2023).

37. Matreyek, K. A. *et al.* Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* **50**, 874–882 (2018).

38. Tareen, A. *et al.* MAVE-NN: learning genotype-phenotype maps from multiplex assays of variant effect. *Genome Biol.* **23**, 1–27 (2022).

39. Klein, F. *et al.* Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell* **153**, 126–138 (2013).

40. Moulana, A. *et al.* The landscape of antibody binding affinity in SARS-CoV-2 Omicron BA.1 evolution. *eLife* **12**, e83442 (2023).

41. Bowley, D. R., Jones, T. M., Burton, D. R. & Lerner, R. A. Libraries against libraries for combinatorial selection of replicating antigen–antibody pairs. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1380–1385 (2009).

42. Dobson, C. S. *et al.* Antigen identification and high-throughput interaction mapping by reprogramming viral entry. *Nat. Methods* **19**, 449–460 (2022).

43. Yu, B. *et al.* Engineered cell entry links receptor biology with single-cell genomics. *Cell* **185**, 4904-4920.e22 (2022).

44. Younger, D., Berger, S., Baker, D. & Klavins, E. High-throughput characterization of protein–protein interactions by reprogramming yeast mating. *Proc. Natl. Acad. Sci.* **114**, 12166–12171 (2017).

45. Li, J. *et al.* Fine antigenic variation within H5N1 influenza virus hemagglutinin's antigenic sites defined by yeast cell surface display. *Eur. J. Immunol.* **39**, 3498–3510 (2009).

46. Mata-Fink, J. *et al.* Rapid conformational epitope mapping of anti-gp120 antibodies with a designed mutant panel displayed on yeast. *J. Mol. Biol.* **425**, 444–456 (2013).

47. Starr, T. N. *et al.* Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295-1310.e20 (2020).

48. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**, 755–768 (2006).

593 49. Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C.-M. An improved yeast transformation method
594 for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155–
595 159 (2010).
596 50. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular
597 biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
598 51. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-
599 End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
600 52. Ouyang, W. O. *et al.* Probing the biophysical constraints of SARS-CoV-2 spike N-terminal
601 domain using deep mutational scanning. *Sci. Adv.* **8**, eadd7221 (2022).

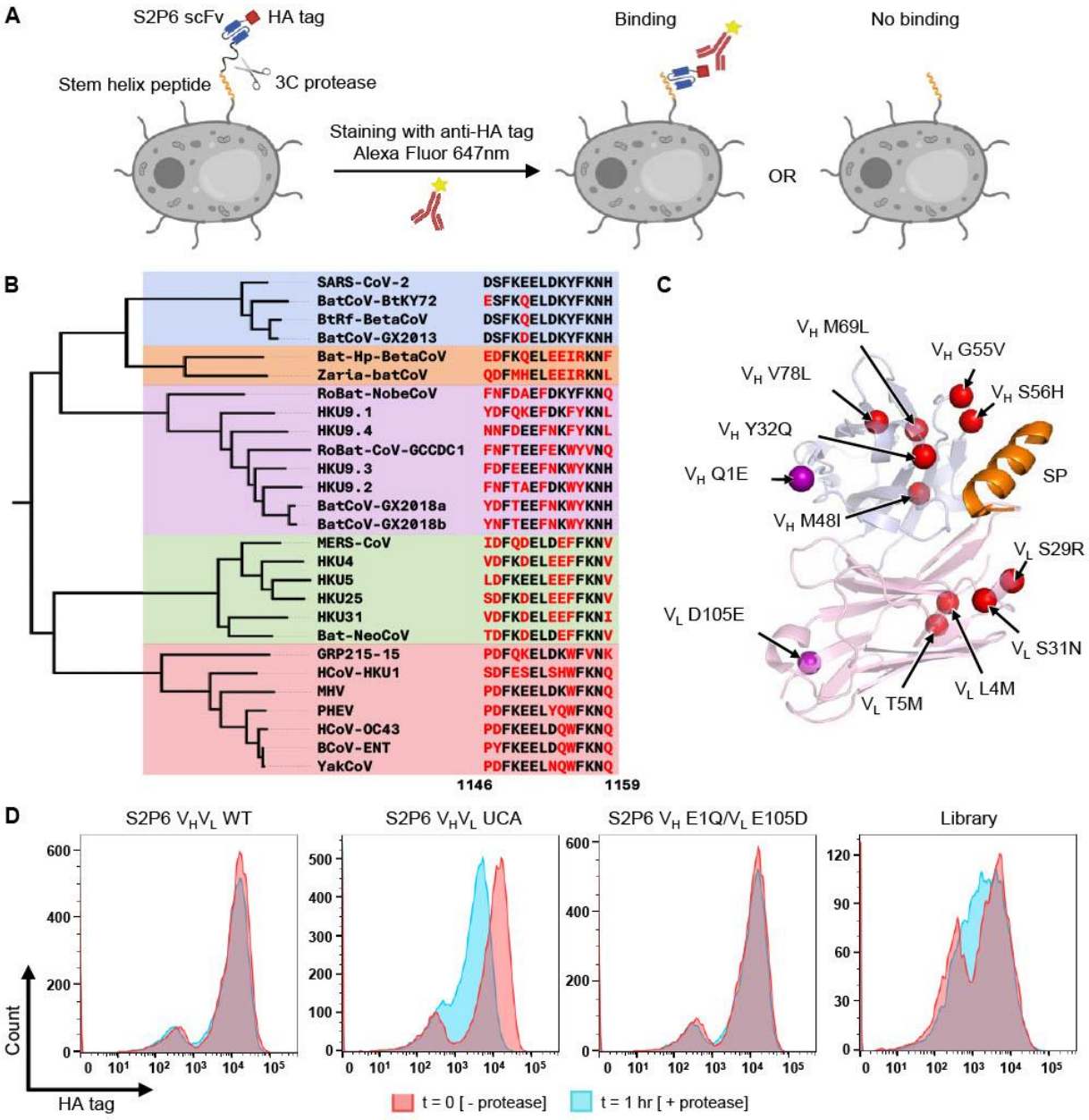
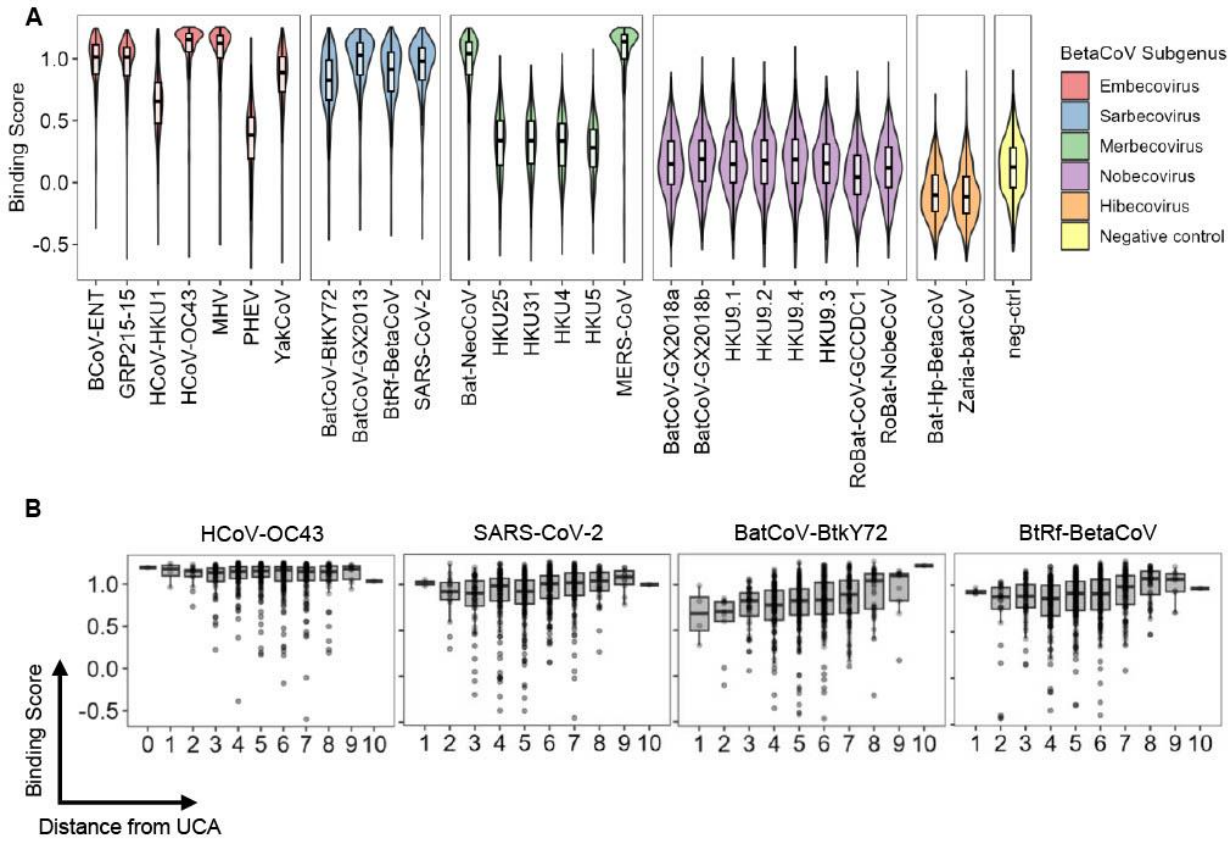


Figure 1. Overview of library-on-library experimental design. (A) Each yeast cell displayed a S2P6 variant in single-chain variable fragment (scFv) format and a stem helix peptide variant with a 3C protease cleavage site in between. After treatment with 3C protease and staining with anti-HA tag Alexa Flour 647 nm, S2P6-SP binding could be assessed using flow cytometry. **(B)** A phylogenetic tree was constructed using the spike sequences of the β-CoV strains included in this study. Sequences of their spike stem helix peptides (residues 1146-1159, SARS-CoV-2 numbering) are shown on the right. Box color denotes β-CoV subgenus: sarbecovirus (blue),

610 hibecovirus (orange), nobecovirus (purple), merbecovirus (green), and embecovirus (red). **(C)**
611 The structure of S2P6 Fab in complex with SARS-CoV-2 stem helix peptide (orange) is shown
612 (PDB 7RNJ)[10], with heavy chain variable domain (V_H) in light blue and light chain variable domain
613 (V_L) in light pink. Somatic hypermutations included in our library design are shown in red and
614 those excluded are in purple. **(D)** Flow cytometry was performed to analyze the HA tag signal
615 (Alexa Flour 647 nm) of yeast cells displaying S2P6 WT, S2P6 UCA, and S2P6 V_H E1Q/V_L E105D
616 with SARS-CoV-2 stem helix peptide, as well as the combination variant library. “- protease” and
617 “+ protease” indicate before and after protease treatment, respectively.



618 **Figure 2. Analysis of S2P6 binding landscapes across stem helix peptides from all β-CoV**
619 **subgenus. (A)** Average binding scores of S2P6 variants to the indicated stem helix peptide
620 variants are shown as a violin plot and a boxplot. SARS-CoV-2 stem helix peptide with two alanine
621 mutations introduced (DSAKEALDKYFKNH), which has previously been shown to abolish binding
622 to S2P6[10], was used as a negative control. **(B)** Binding scores of S2P6 variants with different
623 numbers of somatic hypermutations (i.e. distance from UCA) to HCoV-OC43, SARS-CoV-2,
624 BatCoV-BtkY72, and BtRf-BetaCoV stem helix peptides are shown. Each data point represents
625 the one S2P6 variant. For the boxplot, the middle horizontal line represents the median. The lower
626 and upper hinges represent the first and third quartiles, respectively. The upper whisker extends
627 to the highest data point within a 1.5× inter-quartile range (IQR) of the third quartile, whereas the
628 lower whisker extends to the lowest data point within a 1.5× IQR of the first quartile.

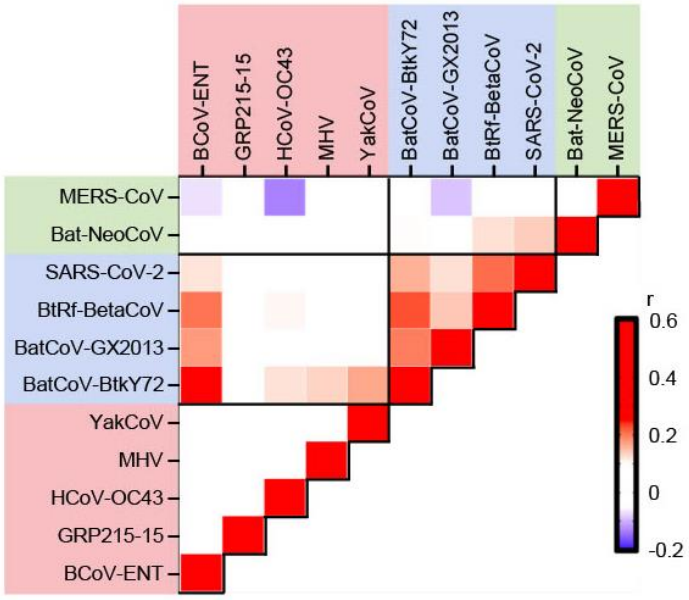
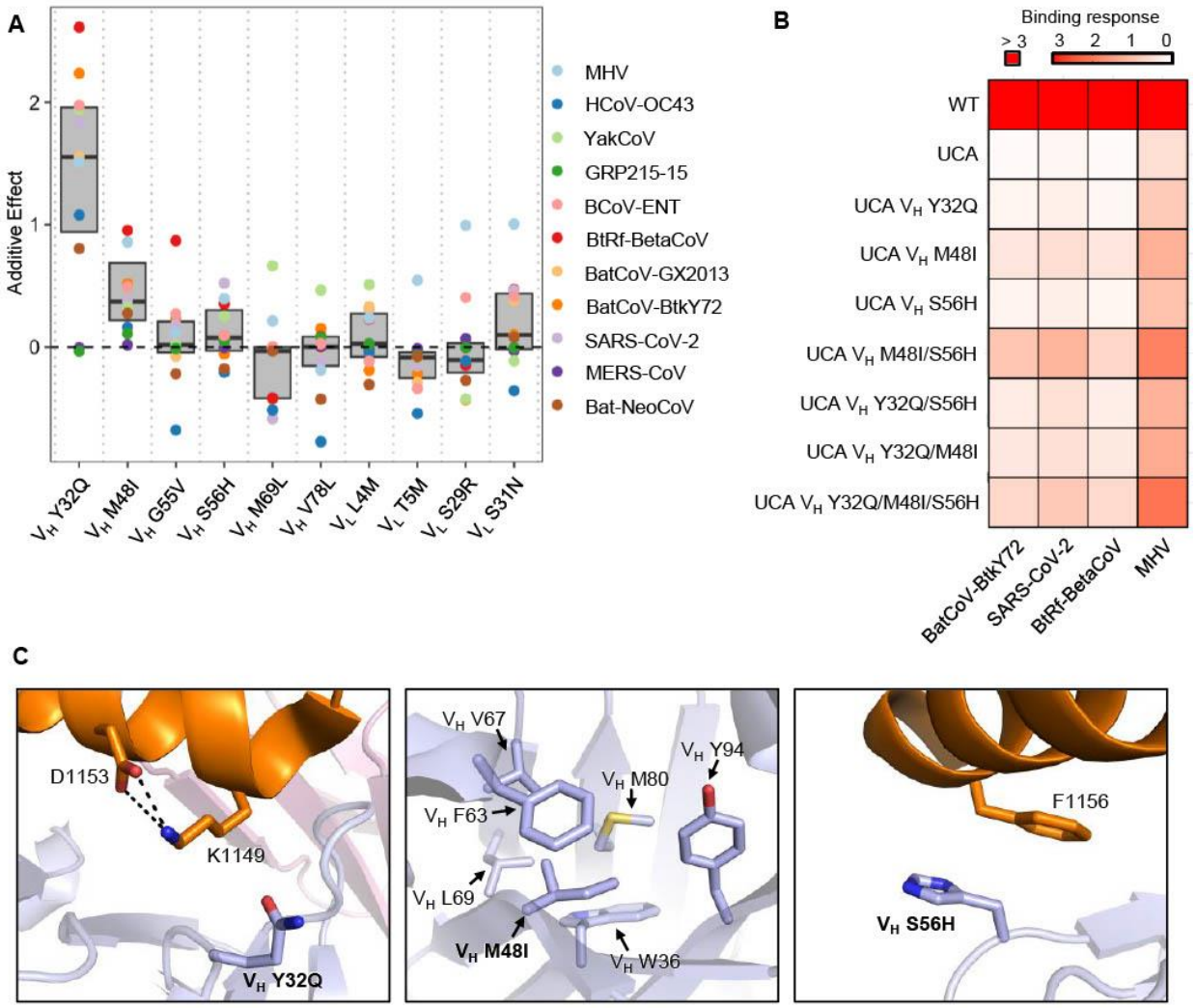


Figure 3. Correlation of S2P6 binding landscapes between stem helix peptides. Pairwise Pearson correlations of binding scores between stem helix peptides are shown. Only those stem helix peptides with an average binding score >0.75 are analyzed here. Red indicates positive correlation coefficient, whereas blue indicates negative correlation coefficient. Diagonal represents the Pearson correlation coefficients between experimental replicates of the indicated stem helix peptides. Shading of axis labels indicates β-CoV subgenus: merbecovirus (green), sarbecovirus (blue), and embecovirus (red).



637 **Figure 4. Key somatic hypermutations for breadth expansion. (A)** The additive effects of the
638 indicated somatic hypermutations on binding to different stem helix peptides are shown. Only
639 those stem helix peptides with an average binding score >0.75 are included here. **(B)** Response
640 levels (nm) of S2P6 WT, UCA, and UCA mutant Fabs binding to BatCoV-BtkY72, SARS-CoV-2,
641 BtRf-BetaCoV, and MHV stem helix peptides were measured using biolayer interferometry. **(C)**
642 Structural analysis of somatic hypermutations V_H Y32Q, V_H M48I, and V_H S56H is shown. SARS-
643 CoV-2 stem helix peptide and S2P6 V_H are shown in orange and blue, respectively (PDB: 7RNJ)[10].