Title
Antibody engineering for the development of enhanced biochemical tools, antivirals, and immunotherapies

Author
Bracken, Colton

Publication Date
2022

Peer reviewed|Thesis/dissertation

Antibody engineering for the development of enhanced biochemical tools, antivirals, and immunotherapies

by
Colton Bracken

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
James Wells
5F2F4D1A06164C2...

James Wells
Chair

DocuSigned by:
William DeGrado
[illegible]...

William DeGrado

DocuSigned by:
Aashish Manglik
4C1E8A184D2E493...

Aashish Manglik

Committee Members

Acknowledgements

There are several factors that made my graduate experience at UCSF remarkable. Chief among them, is the people. The extraordinary mentors, colleagues, friends, and acquaintances have enriched my time here at UCSF with so much joy. For that, I am truly grateful.

Foremost, I want to thank my advisor Jim Wells. Your support and trust have been unconditional. You have shown me how to think outside the box, to be adventurous, and bold in the pursuit of scientific achievement. I am humbled to be your trainee and friend.

The work herein is collaborative and would not be possible without my many collaborators and mentors. Specifically, I want to thank Alex Martinko, and Xin Zhou for teaching me the nuances of phage display, and library construction. Duy Nguyen for helping me persevere in the face of failed projects. Jamie Byrnes for being an exemplar of a tenacious scientist. You are a beacon of positivity in our lab, and I am fortunate to have worked with you and learned from you. I also want to thank the many members of the Wells Lab. Your support has heightened the highs and the smoothed the lows. I want to thank Adam Cotton – for your constant friendship as we've climbed mountains, run hundreds of miles, and brain-stormed startups together; Paige Solomon – my science confidant, your humor, memes, and endless phone calls have kept me afloat during graduate school.

Finally, I want to thank my family, Mark Bracken, Celina Bracken, and Lauren Bracken. Without your constant support I would not have pursued higher education. Mark, your infectious curiosity is the reason I'm a scientist. You taught me to question everything and to explore. Celina, your kindness, and generosity are limitless. You are the driving force that has helped me along the way. Lauren, you remind me not to take life to seriously, your adventurous spirit is inspirational.

Contributions

Chapter 1 of this thesis is a reprint of the material as it appears in:

Zhou, X. X., **Bracken, C. J.**, Zhang, K., Zhou, J., Mou, Y., Wang, L., ... & Wells, J. A. (2020). Targeting Phosphotyrosine in Native Proteins with Conditional, Bispecific Antibody Traps. *Journal of the American Chemical Society*, *142*(41), 17703-17713.

Chapter 2 of this thesis is a reprint of the material as it appears in:

Bracken, C.J., Lim, S.A., Solomon, P. *et al*. Bi-paratopic and multivalent VH domains block ACE2 binding and neutralize SARS-CoV-2. *Nat Chem Biol* **17,** 113–121 (2021).

Chapter 3 of this thesis was completed in collaboration with James Byrnes Ph.D., Xin Zhou Ph.D., Emilo Ramos M.D., Ph.D.; Arun Witta M.D., Ph.D.; Chan-I Chung Ph.D, and Xiaokun Shu Ph.D.

Antibody engineering for the development of enhanced biochemical tools, antivirals, and immunotherapies

By

Colton J. Bracken

Abstract

Antibodies and antibody-like molecules are broadly used as biochemical reagents and therapeutics. They are highly valued as probes that can distinguish between protein targets with extraordinary molecular detail. A deep understanding of natural antibody structure and function has enabled the development of fully synthetic human antibody libraries for *in vitro* display. These synthetic libraries recapitulate the sophistication of molecular recognition in antibody complementarity determining regions (CDRs) and *in vitro* display permits exquisite control over selection conditions. Thus, protein engineers can tailor-make antibodies to bind challenging and non-conventional targets.

Here, we describe three examples of protein engineering that leverage synthetic antibody libraries and *in vitro* display technologies to generate binders to novel epitopes. In Chapter 1, we utilize recombinant antibody pairs to target the post-translational modification (PTM) phosphotyrosine (pY) in folded protein epitopes. In Chapter 2, we develop and deploy a single-domain antibody (sdAb) library to rapidly identify inhibitors to SARS-CoV-2 viral entry. In Chapter 3, we utilize chemo-epitope specific sdAbs to create small-molecule-dependent switches and further engineer them to control cellular therapies.

Table of Contents

List of Figures

List of Tables

Chapter 1

Targeting Phosphotyrosine in Native Proteins with Conditional, Bispecific Antibody Traps

1.1 Abstract

Engineering sequence-specific antibodies (Abs) against phosphotyrosine (pY) motifs embedded in folded polypeptides remains highly challenging because of the stringent requirement for simultaneous recognition of the pY motif and the surrounding folded protein epitope. Here, we present a method named phosphotyrosine Targeting by Recombinant Ab Pair, or pY-TRAP, for in vitro engineering of binders for native pY proteins. Specifically, we create the pY protein by unnatural amino acid misincorporation, mutagenize a universal pY-binding Ab to create a first binder B1 for the pY motif on the pY protein, and then select against the B1−pY protein complex for a second binder B2 that recognizes the composite epitope of B1 and the pY-containing protein complex. We applied pY-TRAP to create highly specific binders to folded Ub-pY59, a rarely studied Ub phosphoform exclusively observed in cancerous tissues, and ZAP70-pY248, a kinase phosphoform regulated in feedback signaling pathways in T cells. The pY-TRAPs do not have detectable binding to wild-type proteins or to other pY peptides or proteins tested. This pY-TRAP approach serves as a generalizable method for engineering sequence-specific Ab binders to native pY proteins.

2.2 Introduction

Protein tyrosine phosphorylation (pY) regulates numerous cellular functions in eukaryotes[1,2]. Recent advances in pY-peptide proteomics have allowed the identification of thousands of pY modifications especially enabled by broadly specific pY binding antibodies (Abs) that do not depend on flanking peptide sequence[3,4]. Crystal structures show that these Abs exquisitely bind the pY residue without contacting neighboring side chains[5]. In addition to generic pY binders, there is a need for specific pY binding probes to follow these events in complex cellular settings and tissues[6,7]. Most commercial phospho-specific Abs were generated using animal immunization with the disordered phosphopeptides[8]. However, this method is low-throughput and expensive, often generates low-affinity, low-specificity, and nonrenewable reagents, and does not apply to nonimmunogenic antigens[7,8].

In vitro display methods based on phage or yeast display have provided powerful renewable and nonanimal derived alternatives for next-generation pY binders with higher specificity and affinity[6,9,10]. For example, several laboratories have reported the engineering of Src Homology 2 (SH2) domains as anti-pY peptide binders to linear peptides with K_D values in the mid- to high-nM range[11,12]. We have previously generated phage libraries derived from a peptide binding Ab that permitted selection of antiphosphoserine and antiphosphothreonine Abs that bind linear peptides with similar affinities and specificities[7]. Higher-affinity pY binders with K_D values in the low-nM range have been identified using an approach called pY-clamps[13]. This utilized a circularly permuted Grb2 SH2 domain linked to an evolvable fibronectin type III (FN3) domain. This biparatopic approach utilized a class of linear pY antigens having a specific peptide binding motif governed by the natural Grb2 SH2 domain.

While linear pY signaling motifs are common in nature, pY sites are also commonly embedded in three-dimensional structural domains[14]. A systematic analysis of phosphorylation sites banked in the mtcPTM database found that pY modifications do not occur more frequently in loops than α-helices or β-strands[15]. Bioinformatics studies presented here further confirm the high frequency of pY modifications in folded epitopes. However, generating highly specific and high-affinity binders to pY sites in tertiary folded protein domains presents significant challenges. While producing a pY peptide immunogen is readily achieved by peptide synthesis, until recently it has been challenging to generate site-specific pY proteins for selections. Furthermore, the recombinant selection approaches using natural motif-based pY binders are well-suited for linear pY epitopes, but these are not easily re-engineered for tertiary folds due to steric hindrance[11-13].

Here, we address the challenge of making pY binders in two folded protein domains, ubiquitin (Ub) and Zeta-chain-associated protein kinase 70 (ZAP70), using a protein engineering platform we call pY Targeting by Recombinant Ab Pairs, or pY-TRAP. In this method, we first generate the site-specifically modified pY protein using a recently described unnatural amino acid misincorporation method[16]. We then identify a first modest affinity binder, B1, that covers the pY modification and the nearby amino acids by selection from a generic 4G10 anti-pY antibody library. We next identify a second, conditional binder, B2, that only binds to the folded pY antigen/B1 complex at a composite site using a Fab-phage library geared toward protein antigens. The identified pY-TRAP binders have high affinity and specificity to the pY proteins.

2.3 Results

A Large Proportion of pY Modifications Occur within a Rigid Secondary Structure

pY-containing sites are often in disordered regions or loops creating linear epitopes for binding partners such as SH2, which are widespread mediators of cell signaling. However, pY modifications can also be embedded in nonlinear, three-dimensional protein structures for which the functions are far less understood (Figure 1.1A). We sought to determine the prevalence of pY modifications that occurred within different secondary structures. The pY-containing sites were retrieved from PhosphoSitePlus (phosphosite.org), a broadly used database that provides comprehensive information on protein PTMs (Figure S1A). In total, 6063 pY sites in 2076 human proteins were identified, along with full-length sequence information from the UniProt database[4,17]. 527 of the proteins (covering 1501 pY sequences) have available three-dimensional structures in the PDB database for the parental, nonphosphorylated form. We used a standard Dictionary of Secondary Structure of Proteins (DSSP) algorithm to extract secondary structure information from the 1501 sequences from the PDB structures (Figure 1.1B)[18]. We found that pY modifications occur in the helix and sheet more than 50% of the time, and this increased to 70% when centered on the exact tyrosine residue. We broadened the analysis to all 2076 pY-containing proteins to predict the secondary structure from the sequence only using the Garnier–Osguthorpe–Robson (GOR) algorithm (Figure 1.1C) and obtained a similar result[19]. As controls, the same algorithms were applied to randomly selected sequences in the studied proteins to determine the average secondary structure, which showed that roughly 55% are helix and sheet (Figure 1.S1B,C). These results highlighted that a large proportion of tyrosine phosphorylation modifications occur in folded epitopes. Developing Abs targeting these three-dimensional pY sites is critical for studying their biological functions.

Generation of Recombinant Biotinylated/Phosphorylated Ubiquitin

Our first native pY protein targeted a largely uncharacterized pY site at position 59 in human Ub. While serine/threonine phosphorylation of Ub is well-known to modulate protein ubiquitination, little is known about how phosphorylation of Y59 impacts Ub function. Ub-pY59 is observed almost exclusively in cancerous tissues by proteomics studies, but its biological relevance remains unclear[16, 20-22].

Methods to prepare proteins with native pY introduced site-specifically have not been available until recently[16, 23]. We generated natively folded Ub-pY59 using an unnatural amino acid (Uaa) misincorporation strategy with a biotinylated Avi-tag for immobilization during phage selection (Figure 1.S2A,B)[16, 24]. Protein biotinylation and pY generation via Uaa deprotection were confirmed by SDS-PAGE gel and liquid chromatography–mass spectrometry (LC–MS) (Figure 1.S2C,D).

Engineering a Generic Anti-pY Ab for Binding to the pY Moiety in Ub-pY59

The overall strategy for pY-TRAPs is shown in Figure 1.2A. To engineer the first pY binder (B1), we started with the commercially available pan-specific recombinant anti-pY Ab, 4G10 [25]. We previously grafted the CDRs onto the highly stable and high expressing trastuzumab scaffold known as 4D5 ($4G10^{4D5}$) and determined the structure (Figure 1.2B)[5]. This antibody fully covers the pY residue, and we found that it bound weakly to Ub-pY59 (K_D of 1.8 μM) as determined by biolayer interferometry (BLI) experiments (Tables 1.S1 and 1.S2). As a specificity control, we tested a generic pY peptide GGG-pY-GGG, with a highly accessible pY and without side chains that may interfere with binding. Indeed, the generic peptide binds to the parental

4G10^{4D5} (K_D = 297 nM) about 5-fold tighter than does Ub-pY59 to the parental (K_D of 1.8 μM). We believe that this reflects the focused contacts with the pY peptide and steric issues for Ub-pY59.

We embarked on further optimization of the 4G10^{4D5} and displayed it on phage. We first tested phage selection mutants of CDR L3 which is in closest contact with the pY to optimize pY recognition and specificity of Ub-pY59 [5]. We identified one mutant L3, 4G10-G4^{4D5}, that was about 5-fold improved in affinity (K_D of 357 nM) relative to the parental 4G10^{4D5} (Figure 1.S3A,B and Table 1.S1). However, the 4G10-G4^{4D5} Fab also bound 2-fold tighter (K_D = 160 nM) to the generic peptide, GGG-pY-GGG, suggesting that the enhanced affinity was partially due to strengthened binding specifically to the pY group[5]. Next, we used a sequential CDR-walking approach to further improve the binding affinity to Ub-pY59 (Figure 1.2B,D, Figure 1.S3C). We first created a phage library with mutations in L1/H1 and another library with mutations in H2 (Lib-L1-H1, Lib-H2, Table 1.S3)[26]. We immobilized the biotinylated Ub-pY59 antigen on streptavidin (SA) magnetic beads and used it for phage selections using a catch and release selection strategy (Figure 1.2C)[5]. Interestingly, the sequences of the selected phage clones showed that the last three residues in H2 are highly conserved, suggesting that these amino acids may play a critical role for binding to the pY group (Figure 1.S3D,E). Next, we combined the mutations from the Lib-L1-H1 and Lib-H2 selections and generated an H3 library based on this new mutant. In addition to varying the amino acid compositions, the length of H3 was also engineered (Table 1.S3). Usually CDR-H3 plays the dominant role among the six CDRs[27,28]. However, in this case, none of the H3 mutants we generated showed strengthened binding to Ub-pY59, suggesting that H3 does not play a dominant role in the interaction of 4G10^{4D5} with Ub-pY59 (Figure 1.S3C). The tightest-binding anti-Ub-pY59 B1 has a K_D of 42 nM to Ub-pY59 and 127 nM to the generic

control peptide (Figure 1.2E, Table 1.S1). The 8.5-fold increase in binding affinity from 4G10-G4^{4D5} to B1 is specific to Ub-pY59 but not to the GGG-pY-GGG peptide, indicating that the gained affinity toward Ub-pY59 resulted from interactions with tertiary sequences surrounding pY and not the pY residue itself.

To increase the efficiency for B1 engineering, based on the knowledge learned from these CDR-walking experiments, we generated a phage library by randomizing sequences in CDR L1, H1, and H2 using 4G10-G4^{4D5} as the parental sequence (Lib-L1-H1-H2, Figure 1.2F, Figure 1.S3A and Table 1.S3). Using this single-step library, we identified two Abs, B1–2 and B1–3, that interacted with Ub-pY59 with K_D values of 63 or 79 nM, respectively, which is ~5-fold tighter than the parental clone and only slightly weaker than the variant identified using the more laborious sequential CDR-walking approach (K_D = 42 nM) (Figure 1.2G, Table 1.S1). Although these B1 selectants bind significantly tighter to Ub-pY59 than the parental 4G10^{4D5} Fab, they also bind only 2-fold weaker to the GGG-pY-GGG control peptide (K_D of 127 nM), indicating that B1 would likely recognize other pY modifications.

Engineering B2 for Binding to the Ub-pY59/B1 Complex

To further improve binding specificity and affinity, we sought to identify a second binder (B2) that conditionally binds the complex of B1/Ub-pY59, but not to B1 nor Ub-pY59 alone, nor to the B1/GGG-pY-GGG complex (Figures 1.2A and 1.3A). To select for such a conditional B2 Fab, we utilized a synthetic Fab-phage library extensively used for selection of globular protein antigens[29]. We first removed Fabs that bind to Ub-pY59 using a negative selection by incubating with immobilized Ub-pY59 on SA beads (Figure 1.3A). The B1/Ub-pY59 complex positive selection bait was generated by mixing SA-immobilized Ub-pY59 with 5 μM of B1-Fab and

incubating for 30 min at room temperature (RT). This B1/Ub-pY59 complex was then incubated with the Ub-pY59-cleared phage library to pull down complex-specific binders. The B1-Fab (5 μM) was added both to the phage library/beads mixture as well as to the wash buffers to maintain the formation of the B1/Ub-pY59 complex. Having excess amounts of B1-Fab in solution also prevented enrichment of binders that interact with B1 alone on the SA beads. Phage titer experiments showed that binders were specific to the B1/Ub-pY59 complex, as they enriched as a function of round of selection (Figure 1.3B).

Next, we characterized the binding of 96 phage clones to the B1/Ub-pY59 complex or Ub-pY59 alone using a phage enzyme-linked immunosorbent assay (ELISA). A group of Fab phage clones were identified that selectively interact with the complex but not Ub-pY59 alone (Figure 3C). These Fab-phage clones were recombinantly expressed as Fabs or IgGs for further characterization. A sequential BLI experiment was performed to identify the tightest binding B2 against the B1/Ub-pY59 complex (Table 1.S2). The B1-Fab bound to both Ub-pY59 and GGG-pY-GGG, but remarkably the B2 only interacted with the B1/Ub-pY59 complex (K_D = 5 nM for Fab and 0.6 nM for IgG) and not detectably to the B1/GGG-pY-GGG complex (Figure 1.3D). In addition, neither B1-Fab nor B2-Fab recognized the nonphosphorylated Ub-Y59 (WT) (Figure 1.3D). These results demonstrate that the B2 TRAP binder is highly selective to the B1/Ub-pY59 complex. In effect, the pY-TRAP is a conditional biparatopic binder that depends on B1 binding the pY protein before B2 can engage it and the native protein.

In the phage ELISA experiment, we also observed a number of nonconditional clones that interacted with both the Ub-pY59 and the B1/Ub-pY59 complex (Figure 1.3C). The enrichment of these binders was likely due to the incomplete negative clearance of the library with Ub-pY59. One of these clones, designated as B_{WT}, was expressed and characterized in BLI experiments.

Unlike B2, B_{WT} not only bound to the B1/Ub-pY59 complex but also bound to unmodified Ub-Y59 (WT) (Figure 1.S4).

We further interrogated the specificity of the anti-Ub-pY59 TRAP binders by testing the binders against a panel of pY protein or peptide antigens in an ELISA assay (Figure 1.3E). The pY proteins or peptides (100 or 10 nM) were immobilized on an ELISA plate, followed by incubation with a mixture of anti-Ub-pY59 B1-Fab and B2-IgG. An anti-IgG-HRP secondary Ab that binds to B2-IgG was used to detect binding of B2-IgG to the pY-antigen in complex with B1-Fab. We observed the ELISA signal only with the Ub-pY59 antigen, which further highlighted the specificity and utility of the identified TRAP binders.

The engineered anti-Ub-pY59 B1 binder has a binding affinity of 42 nM to the antigen. Of note, when we tried to select for a B2 binder against the parental 4G10-G4^{4D5}/Ub-pY59 complex that interacts with almost 10-fold weaker affinity (K_D = 357 nM), no enrichment in the phage titer measurement was observed. This suggests that a moderate binding affinity between B1 and pY antigen is important for the successful enrichment of B2 binders by phage display.

Engineering and Characterization of B1–B2 Fusions

We next explored how a B1 and B2 fusion performs as an anti-Ub-pY59 binder. B1 was converted to a single-chain variable fragment (B1-scFv) and fused to the C terminus of B2-Fab heavy chain (HC) with a 26-aa linker (Methods section, Figure 4A). The scFv-HC plasmid was coexpressed with the B2-Fab light chain (LC) plasmid to generate the B1–B2 fusion. We found that the B1–B2 fusion bound Ub-pY59 with a very high affinity (K_D = 0.5 nM), while it interacted with GGG-pY-GGG with a K_D of 133 nM. We further tuned the linker length and found that a 14-residue linker resulted in the greatest difference in K_Ds for Ub-pY59 and GGG-pY-GGG (0.5 vs

147 nM) (Figure 1.4A,B, Table 1.S1). This fusion protein does not interact with Ub-Y59(WT), Y59A, or Y59E variants showing strong dependence of pY59 (Figure 1.S5A).

To explore the molecular mechanism driving the enhanced affinity of the B1–B2 fusion to Ub-pY59, we used negative stain electron-microscopy (NS-EM) to study the structure of the protein complex. The complex of the 14-residue B1-scFv/B2-Fab fusion and Ub-pY59 was prepared by mixing equal molar amounts of the two proteins, which were then purified by size exclusion high-performance liquid chromatography (HPLC) (Figure 1.4C). The peak of the protein complex was collected, and a fraction was run on SDS-PAGE gel to validate the formation of the equal stoichiometric complex (Figure 1.S5B). However, the retention time of the complex on HPLC indicated that the size of the B1–B2/Ub-pY59 complex was ~150 kDa, which is twice the size of a 1:1 complex (~88 kDa) (Figure 1.4C). Furthermore, while NS-EM imaging showed that the B1-scFv/B2-Fab fusion has an expected scFv-Fab monomeric structure (Figure 1.4D,E), NS-EM 2D averages of the complex showed that it is much bigger than B1-scFv alone. The structure clearly contains two "donut-shape" Fab structures in each molecule (Figure 1.4F). Together, these results indicated that B1–B2 and Ub-pY59 formed a 2:2 complex.

To further study the 2:2 complex, we modeled the structures in PyMOL and found that the B1 and B2 could easily accommodate domain swapping within a single B1–B2 fusion polypeptide and two different Ub-pY59 molecules to produce a 2:2 complex (Figure 1.4G,H). This 2 on 2 cooperative binding model results in intermolecular avidity contributing to higher affinity (Figure 1.4B, Table 1.S1). Furthermore, based on this structural model, we estimated that a long linker (>100 Å, approximately 50 residues) would be necessary for B1 and B2 intramolecular association to simultaneously bind a single Ub-pY59 molecule and form a 1:1 complex (Figure 1.S5C)[30]. We tried to express a B1-scFv/B2-Fab fusion with longer linkers but obtained very little

protein with a linker as long as 40 aa. This result suggested that a long linker fusion is not an ideal molecular format (Figure 1.S5D). Further studies will explore if a differing fusion design, such as reversing the order of B1/B2 fusion, using two scFvs instead of a scFv and a Fab, making a bispecific Fc linked parallel construct, or fusing each of B1 and B2 to an interaction domain, would lead to a binder that can form a 1:1 complex with Ub-pY59.

The B1–B2 fusion to Ub-pY59 represents one of the tightest synthetic binders (K_D = 0.5 nM) to PTMs to our knowledge (Figure 1.4I). It has great selectivity against the GGG-pY-GGG peptide which is a very stringent off-target control for binding generic pY because of its high accessibility (K_D = 147 nM). As an additional control that binding depends on surrounding protein structure, we incorporated pY using the described Uaa approach into E64 which is highly exposed in Ub, to produce Ub-pY64. The anti-Ub-pY59 B1–B2 fusion did not show any detectable binding to Ub-pY64 further confirming its high specificity for Ub-pY59 (Figure 1.S5E).

Phosphoproteomics with 4G10^{4D5} Variants Reveal a Large pY-Target Landscape for pY-TRAPs

To explore the scope of targetable pY antigens using the 4G10^{4D5} phage library, we performed phosphoproteomics on lysates of Jurkat cells using WT 4G10, 4G10-G6^{4D5}, and an SH2 variant, called Src-Superbinder, another broadly used pan-specific anti-pY binding domain[11]. Briefly, pY peptides were first immunoprecipitated from trypsin-treated cell lysates (Figure 1.5A) with Ni-NTA resin-immobilized binders, followed by TiO_2 enrichment. Using a standard pY peptide enrichment workflow, we identified 4824 pY-sequences with 4G10, 4775 with 4G10-G6^{4D5}, and 4446 with Src-Superbinder (Figure 1.5A,B, Appendix Table 1.S5).

To further explore the accessibility of pY on folded proteins, we repeated this experiment with an additional step of immunoprecipitation from Jurkat whole cell lysates (Figure 5A) prior to trypsinization to capture native pY proteins. Using the pY protein+peptide enrichment workflow, we identified 2290 pY-sequences with 4G10, 2289 with 4G10-G64D5, and 1746 with Src-Superbinder (Figure 1.5C, Appendix Table 1.S4). The ~50% fewer pY sequences identified using the protein+peptide enrichment workflow confirmed that a significant proportion of native pY motifs are not readily accessible to generic pY binders due to steric hindrance of the folded protein epitope. However, the thousands of native pY sequences identified from this proteomics experiment indicated that 4G10 has substantial initial binding affinities to a large number of candidate proteins suited for pY-TRAP engineering. The Src-Superbinder enriched a fewer set of pY sequences. There was not complete overlap suggesting that both types of generic pY binders could allow a better pY sequence coverage[31]. These data also suggested candidates for pY-TRAPs using SH2 domain libraries to generate potential B1 binders.

Generalizing the pY-TRAP Method to ZAP70-pY248

It is not uncommon for proteins to have multiple phosphotyrosine sites, and we wished to test a protein in this class both for robustness of pY misincorporation and pY-TRAP generation. Within our Jurkat cell proteomics data set, we identified 11 phosphorylated sites in the ZAP70 kinase known to play a role in T cell activation and feedback signaling mechanisms (Appendix Tables 1.S4 and 1.S5)[32, 33]. Three of these sites (Y178, Y221, Y248) are within the second SH2 domain (residues 167–259). Most of these sites do not have commercially available pY-Abs, preventing us from gaining a more comprehensive understanding of their functions. Using the described Uaa approach we achieved robust expression of pY221 and pY248 within the SH2

domain but not for pY178 (Figure S6)[16]. This is not surprising as it is known that expression of proteins by unnatural amino acid misincorporation can vary depending on target sequence possibly due to ribosome stalling and other effects.

We subjected the pY221 and pY248 antigens to phage selection using a 4G10-G6^{4D5} Lib-L1-H1-H2 phage library to identify B1 binders (Table 1.S3). An anti-ZAP70-pY248 B1 binder was identified that was ~3-fold tighter (K_D = 21 nM) to ZAP70-pY248 than the parental 4G10-G6^{4D5} (K_D = 56 nM) (Figure 1.5D, Table 1.S1). Separately, we selected for B1s to ZAP70-pY221, but none of the isolated B1s bound significantly tighter. The affinities of B1-Fabs to Ub-pY59 and ZAP70-pY248 were higher (K_D = 42 nM and 21 nM, respectively) than the binding of the B1 for ZAP70-pY221 (K_D = 75 nM).

We further optimized the B2 engineering workflow (Figure 1.S7). Instead of providing a high concentration of B1-Fab in the previous process (Figure 1.2A), we coimmobilized both pY protein and B1 on SA beads to create a high local concentration of the two interacting partners. We hypothesized that this would stabilize the pY protein/B1 complex avoiding the need for high concentrations of B1 in solution. To test this hypothesis, we performed a B2-IgG pulldown experiment with biotinylated Ub-pY59 and biotinylated B1 immobilized on SA beads. Supplementing additional B1-Fab in solution did not result in an increased level of B2 pulldown, suggesting that immobilizing both Ub-pY59 and B1 on the beads had created a stable complex (Figure 1.S7A).

Based on this result, we performed B2 selections for the pY ZAP70 antigen-B1 complexes by immobilizing both biotinylated ZAP70-pY antigens and biotinylated B1 on SA beads (Figure 1.S7B). Briefly, a Fab-phage library was negatively selected against the immobilized ZAP70 pY antigens, followed by another negative selection against immobilized B1 binders. Next, a positive

selection was performed with preincubated, SA-immobilized pY protein/B1 complexes. After four rounds of selection, we identified a B2 binder highly specific to B1/ZAP70-pY248 (Figure 1.5E, K_D = 4 nM). Of note, no complex specific binders were identified for the weaker B1/ZAP70-pY221 complex. It is possible that there is a threshold affinity to select B2s to this pY protein-B1 complex. BLI and ELISA experiments demonstrated that the B2 binder is highly specific to the B1/ZAP70-pY248 complex (Figure 1.5E,F, Figure 1.S8).

1.4 Discussion

Engineering high-affinity recombinant binders using multiple binding arms has been used extensively to increase affinity and specificity: for example, the generation of bispecific Abs where two independently selected Abs bind two separate-linked or membrane bound proteins, or biparatopic Abs where two independent binders with nonoverlapping but neighboring epitopes are linked together on the same protein[34]. The pY-TRAP represents a novel class in that B2 binding was conditional to the presence of the complex of the pY protein and B1. During our B2 selections, we did find independent binders (we called B_{WT}), some of which we could have used as biparatopic constructs but did not as B_{WT} could bind independent of phosphorylation. Phosphotyrosine forms can have a range of stoichiometry, and we deliberately did not want B2 to bind independently to the pY protein or B1 as this would reduce specificity for the complex.

We specifically chose the 4G10[4D5] as a broadly specific B1 scaffold because of its exquisite dependence on pY, ability to bind pY in tertiary epitopes, high level expression in *Escherichia coli*, and single selection library to further enhance potency surrounding the pY residue. Although this starting scaffold will certainly not bind all pY proteins, the pY proteomics here provides a list of about 2300 pY sites (from Jurkat extracts alone) that were pulled down with the 4G10 Ab which

could be possible candidate pY targets for pY-TRAP selections. Another limitation is that not all pY protein domains will be amenable to the unnatural pY incorporation due to large size, acid sensitivity, or other expression issues. Fortunately, both pY sites in Ub, and two of three sites tested in ZAP70, did express. An alternative method to generating pY proteins was recently published using a different Uaa and synthetase which provides another option for generating the pY antigens[23]. Also, not all B2 selections worked, and this may be because the B1–pY protein affinity was not sufficient to present the complex. It is possible that coimmobilizing higher concentrations of B1 and pY protein will mitigate this by enforcing high complex formation. In addition, the current B1 library may not be the most optimal for all pY proteins. Further exploration on how each of the CDRs contribute to binding to a broader panel of pY antigens will promise a further improved, second-generation library for engineering tighter B1 binders.

The pY-TRAP approach enabled us to generate high-affinity and specific binders against native Ub-pY59 and ZAP70-pY248, for which there are no commercially available Ab reagents. We believe that building pY protein binders to tertiary pY epitopes nicely complements and expands approaches for linear pY epitopes using diversified SH2 or PTB domains and pY-Clamps (Figure 5G)[11-13]. Although protein kinase substrate specificity is typically defined by linear consensus motifs, recent evidence suggests that protein-folding creates structure-based motifs for specific kinase recognition[14,15, 35-38]. Consistent with these findings, our bioinformatics analysis revealed that tyrosine occurs very frequently in rigid tertiary structures. pY-TRAPs serve as vital reagents for assigning the biological functions, determining the upstream kinases, and studying signaling regulations of this class of pY modifications.

We envision that this conditional two-binder TRAP method may be generalizable to other PTMs if generic B1 binding scaffolds are developed to PTMs (Figure 5H). In this regard,

recombinant anti-methylated (Me)-lysine Abs that independently bound two neighboring Me-lysine sites on histone H3 were recently reported[39]. Interestingly when the Abs were mixed, they discovered that they bound cooperatively with higher affinity. Additionally, natural domains that recognize a different protein PTM could be used as B1, for example, SH2 variants for sulfotyrosine, and lectins for glycosylation, while B2 domains bind the complex of B1 and PTM-containing antigen[40-43]. Overall, there is a great potential to generalize this TRAP platform to additional protein modifications for making a wide variety of PTM-TRAPs to serve as research agents, biomarkers, or therapeutics.

Acknowledgements:

We gratefully acknowledge members of the Wells lab for helpful discussions and support. We gratefully thank Nicholas Young for help with ZAP70 phosphoprotein generation and characterization. We also valued helpful discussions with Dr. Sachdev Sidhu

1.5 Materials and Methods

Secondary Structure Analysis

Secondary structures for pY sequences were performed using an in-house informatics pipeline written in R. Scripts are available for download from https://github.com/crystaljie/pY_2rd_structure_solvent_accessibility_analysis.git. The protein phosphorylation site information used for this analysis was downloaded from PhosphoSitePlus (Appendix Table S1). Only the pY sequences in human proteins were used for analysis. Based on the full-length protein sequences from Uniprot, the GOR algorithm was applied to predict the secondary structure of the pY sites (6063 pY sites in 2076 proteins) or

randomly picked sequences from these proteins (Appendix Table 1.S2). For the sequences that have PDB structures available (1501 pY sites in 527 proteins), DSSP algorithm was used to obtain the exact secondary structure information for the pY sites or randomly picked sequences from these proteins (Appendix Table 1.S3)[17-19].

Plasmid Construction

Plasmids were constructed by standard molecular biology methods, and complete plasmid sequences are available upon request. The DNA fragment of human Ub or ZAP70 was synthesized by integrated DNA technologies (IDT). The TAG mutation for pY incorporation was introduced by overlap-extension PCR. The antigens were subcloned into the inducible bacterial expression pTak vector with a Tobacco Etch Virus (TEV) protease cleavage site, an AviTag, and a six-consecutive-histidines tag (His-tag) at the C-terminus[16,44]. The plasmid pBK-MmNpYRS encoding the Uaa-specific synthetase was provided by the Lei Wang lab (University of California, San Francisco)[16]. All the Fabs were constructed in a dual-expression vector that expresses the light chain and the heavy chain with the pelB and the stII signal peptides, respectively, for the periplasm expression[5]. A C-terminal 6× His tag was put in the heavy chain for purification. All the IgGs were constructed in the pFUSE-hIgG1 vector (InvivoGen), with the heavy chain genetically fused to the hIgG1-Fc on one vector, and the light chain on a separate copy of the vector. The B1-scFv/B2-Fab fusion was also constructed in the pFUSE-hIgG1 vector (InvivoGen), with B2-Fab light chain on one vector, and B2 Fab heavy chain fused to the N-terminus of B1-scFv on another copy of the vector. The sequences of the linker between the heavy chain and scFv are as follows: 26 aa, GGSGSAGGLNDIFEAQKIEWHESSGS; 19 aa, AGGLNDIFEAQKIEWHEGS; 14 aa, AGGLNDIFEAQKGS; 9 aa, AGGLNDIGS.

Fab Expression

C43 (DE3) Pro+ *E. coli* containing Fab expression vectors were grown in 2xYT at 37 °C to an OD-600 of 0.4–0.8, and then, Fab expression was induced by the addition of 0.5 mM IPTG. Incubation temperature was subsequently reduced to 30 °C, and the cultures were allowed to shake for 16–20 h. Cells were harvested by centrifugation and lysed using B-PER lysis buffer or sonication. The lysate was incubated at 60 °C for 20 min and centrifuged to remove the inclusion body. The Fabs were purified by Ni^{2+}-NTA resin and buffer exchanged in TBS buffer (50 mM Tris, 150 mM NaCl, pH 7.5) for further characterization.

IgG and B1-scFv/B2-Fab Fusion Expression

The IgGs were expressed by cotransfection of the pFuse-light chain and the pFuse-heavy chain-Fc vectors. The B1-scFv/B2-Fab fusion proteins were expressed by cotransfection of the pFUSE vector encoding B2 Fab light chain and another pFUSE vector encoding B2 Fab heavy chain fused to the N-terminus of B1-scFv. Expi293 (Life Technologies) cells were transiently cotransfected with two vectors at a mass ratio of 1:1. The ExpiFectamine 293 transfection kit (Life Technologies) was used for transfections as per manufacturer's instructions. Cells were incubated for 5 days at 37 °C in a 5% CO_2 environment before the supernatants were harvested by centrifugation. The IgGs and the B1-scFv/B2-Fab fusion were purified by Protein A and Protein L affinity chromatography, respectively.

pY Misincorporation

pY was incorporated into Ub and ZAP70 according to a previous protocol[16]. Briefly, BL21 cells were cotransformed with the pTak-antigen plasmids (chloramphenicol resistant) and pBK-MmNpYRS (kanamycin resistant). 35 μg/mL chloramphenicol and 50 μg/mL kanamycin were added to the bacterial culture to select the double transformants. The cells were grown in 2xYT at 37 °C to an OD-600 of 0.4–0.8, and then, 1 mM Uaa 1 (Figure 1.S1) was added. Protein expression was then induced by the addition of 0.5 mM IPTG. The incubation temperature was subsequently reduced to 18 °C, and the cultures were allowed to shake for 16–20 h. Cells were harvested by centrifugation and lysed by sonication. The lysate was centrifuged to remove cell debris and the inclusion body. The proteins were purified by Ni^{2+}-NTA resin and exchanged into TBS buffer.

To cleave the protection group on the Uaa, to a diluted protein solution (0.1–1.0 mg mL^{-1}) in TBS buffer, HCl (4 M) was added to reach a final concentration of 0.4 M. The reactions were incubated at 4 °C for 24–72 h, and the cleavage efficiency was monitored by mass spectrometry on a Xevo G2-XS mass spectrometer (Waters).

In Vitro Biotinylation

The purified proteins were biotinylated on their AviTags using the standard protocol provided by Avidity. Biotinylation was monitored by intact protein mass spectrometry on a Xevo G2-XS mass spectrometer (Waters). The biotinylated proteins were then separated into aliquots, analyzed by SDS-PAGE, snap-frozen, and stored at −80 °C for later use.

Phage Library Construction

The phagemids that encode the $4G10^{4D5}$ gene with the stop codon within the CDRs were used as templates for Kunkel mutagenesis with oligonucleotides designed to correct the stop codons and introduce the designed mutations at each site[7,45]. The library designs are listed in Table 1.S3. The resulting mutagenesis reactions were electroporated into SS320 electrocompetent cells (Lucigen). After 1 h of recovery at 37 °C, the cells were expanded into 500 mL of 2xYT cultures with 50 μg/mL carbenicillin and 10^{10}/mL M13KO7 helper phage. 50 μg/mL kanamycin was added to the culture after 1 h of shaking at 37 °C. The culture was grown for approximately 20 h with 250 rpm shaking at 37 °C. The next day, the cells were pelleted by centrifugation. The phage was precipitated from the supernatant by adding 1/5 volume of 20% PEG8000 and 2.5 M NaCl. The phage library was resuspended in TBS buffer with 50% glycerol and 2 mM EDTA and stored at −20 °C. Template phagemid plasmid, mutagenesis oligos, and a small quantity of library stocks can be shared with the academic community upon request.

Phage Display

All phage selections were done according to previous protocols[5,46]. For B1 selection, the $4G10^{4D5}$ libraries were incubated with biotinylated pY protein immobilized on SA-coated magnetic beads (Promega). Empty beads were used for library clearance and enrichment tests. In total, four rounds of selections were performed with a decreasing concentration of pY-antigen (300, 100, 30, 10 nM). Starting from round 2, the phage library was first enriched by protein L magnetic beads to deplete nondisplayed or truncated Fab phage before each round of the selection. For B2 selection for Ub-pY59, the synthetic phage library was first incubated with pY protein immobilized on SA beads to deplete any binders to the pY protein alone. Subsequently, the cleared library was incubated with immobilized pY-antigen with 5 μM B1-Fab in solution to enrich

binders against the pY protein/B1 complex. In total, four rounds of selections were performed with 500, 500, 500, and 100 nM of pY antigen, respectively. B1-Fab (5 μM) was supplemented to both the selection reaction and the wash buffers.

For B2 selection for ZAP70-pY221 and pY248, the synthetic phage library was first incubated with immobilized pY protein and B1, respectively, to deplete any binders to pY protein or B1 alone. Subsequently, the cleared library was incubated with the B1/pY-antigen complex (with both B1 and pY-antigen immobilized on the SA beads) to enrich binders against the pY protein/B1 complex. In total, four rounds of selections were performed with 500, 500, 500, and 100 nM of pY antigen, respectively.

ELISA

ELISAs were performed according to standard protocols. Briefly, 384-well Maxisorp plates were coated with NeutrAvidin (10 μg/mL) overnight at 4 °C and subsequently blocked with BSA (2% w/v) for 1 h at 20 °C. 10–100 nM pY antigens were captured on the NeutrAvidin-coated wells for 30 min followed by the addition of various concentrations of phage or recombinant antibodies for 30 min. The secondary Abs were either a horseradish peroxidase (HRP)-conjugated anti-M13 phage antibody (Sino Biological) for phage ELISA or an antihuman-IgG antibody (Sigma-Aldrich) for recombinant protein ELISA. The ELISA plates were washed three times after each incubation, and Ab binding was detected by TMB substrate (VWR) and read at 450 nm.

Binding Kinetics Analysis

BLI experiments were performed at room temperature using an Octet RED384 instrument (ForteBio). Biotinylated pY antigens were immobilized to an optically transparent SA biosensor

(ForteBio). Different concentrations of antibodies in kinetics buffer (TBS, pH 7.5, 0.05% Tween-20, 0.2% BSA) were used as the analyte in a 384-well microplate (Greiner Bio-One). Antibody CDR sequences are listed in Table 1.S2. Affinities (K_Ds) were calculated from a global fit (1:1) of the data using the Octet RED384 Data Analysis HT software. Measurements at the single antibody concentration were first used to estimate binding strengths of different selectants, and measurements at a series of antibody concentrations (covering concentrations below and above the estimated K_D) were used to determine K_D values (Table 1.S1). The measurements were performed at least twice (single- or multiconcentration), and representative Octet traces are shown in Figure 1.S9.

Negative Stain Electromicroscopy

The samples B1-scFv/B2-Fab fusion and its complex with Ub-pY59 were negatively stained and observed on a Tecnai T12 microscope and a Tecnai T20 microscope (FEI Company) using the discharged continuous carbon grids as previously described, respectively. Images were acquired at room temperature with a pixel size of 2.21 Å/pixel (T12, operated at 120 kV) or 3.319 Å/pixel (T20, operated at 200 kV) on the level of specimen using a 4K × 4K CCD camera (UltraScan 4000, Gatan Inc.)[47,48]. After all micrographs were visually screened, the contrast transfer function (CTF) was estimated for each micrograph by Gctf [49]. The particle was selected using the Gautomatch (https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch) without template. Individual particles were extracted from the raw images with the 100 × 100 pixel window for T12 images or 80 × 80 pixel window for T20 images and were subjected to 25 cycles of 2D classification with a mask diameter of 200 Å for B1-scFv/B2-Fab or 220 Å for its complex with Ub-pY59 in Relion 3.0

[50]. The output 2D averages after several rounds were analyzed by comparing with the available protein structures utilized in this design.

pY Phosphoproteomics

The phosphoproteomics experiment was performed according to the standard pY phosphoproteomics protocol and manufacture protocols [31]. Briefly, Jurkat cells were pretreated with 0.1 mM freshly activated pervanadate for 15 min at 37 °C. The cells were spun down and washed with prechilled (4 °C) TBS (20 mM Tris, 150 mM NaCl), and lysed in prechilled (4 °C) lysis buffer (0.5% Triton-100, 50 mM Tris, 150 mM NaCl) with phosphatase inhibitor cocktail (Sigma-Aldrich, 1:100) and protease inhibitor cocktail (Sigma-Aldrich, 1:100). After rotating at 4 °C for 15 min, cell lysates were spun at 16000*g* to remove cell debris. 10 nmol his-tagged 4G10, 4G10-G6^{4D5}, or Src-Superbinder was loaded to 100 μL of prewashed Ni^{2+}-NTA resin (Thermo Fisher Scientific) and washed twice with TBS buffer. For phosphoprotein IP, the cell lysates were added to the Ni^{2+}-NTA slurry and rotated for >2 h at 4 °C. For phosphopeptide IP, proteins were first denatured, reduced, alkylated, digested to peptides by MS-compatible, sequencing grade trypsin (Promega, 1:50 w/w), and then added to the Ni^{2+}-NTA resin for pulling down pY peptides. The Ni^{2+}-NTA resin was washed 6 times with TBS buffer and eluted with 100 mM phenyl phosphate in 500 mM ammonium bicarbonate buffer. The elution was desalted using a SOLA column, and phosphopeptides were further enriched by the TiO_2 phosphopeptide enrichment kit (Thermo Fisher Scientific). The elution from TiO_2 enrichment was resuspended in 2% ACN + 0.1% FA and analyzed by a Q Exactive Plus (Thermo Fisher) mass spectrometer.

Immunoprecipitation with Anti-Ub-pY59-TRAP Binders

2.5 μM biotinylated B1-Fab and biotinylated Ub-pY59 were incubated for 30 min at RT. This B1/Ub-pY59 complex or 2.5 μM biotinylated Ub-pY59 only was loaded onto the SA beads. 0 or 5 μM of B1-Fab was supplied in solution. Subsequently, 5 μM B2-IgG was added to the reaction mixtures and incubated for 30 min at RT. The beads were washed with TBS buffer or TBS+5 μM B1-Fab. Proteins were eluted with 0.1 M acidic acid.

1.6 Main Figures

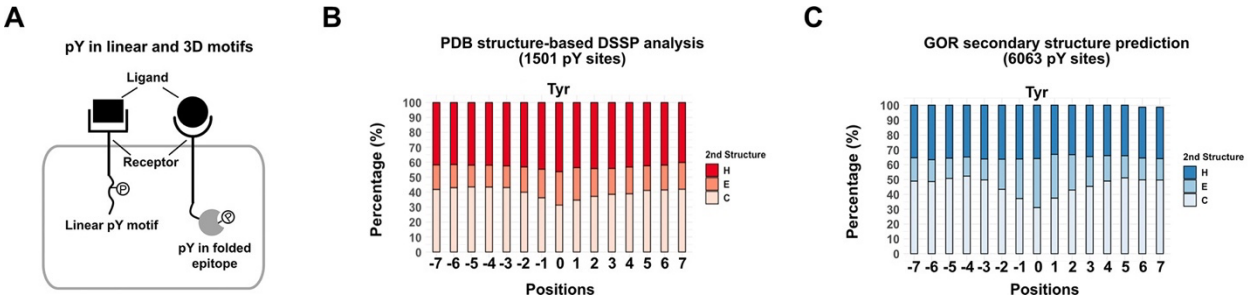


Figure 1.1: Secondary structure analyses of pY sequences suggest that pY epitopes are commonly embedded in 3D epitopes containing a helix or sheet structure. (A) Schematic illustration of linear and three-dimensional (3D) pY motifs in the receptor proteins in a cell. (B) Secondary structure analysis of 1501 pY sites using the DSSP algorithm based on available PDB structures. (C) Secondary structure prediction of 6063 pY sequences using the GOR algorithm. Position 0 is the pY residue; positions –7 to –1 or 1 to 7 are the seven residues upstream or downstream of the pY motif. H, helices (3,4,5-turn helix); E, strand (ß-sheet, ß-bridge); C, loop (bend, turn, coil)

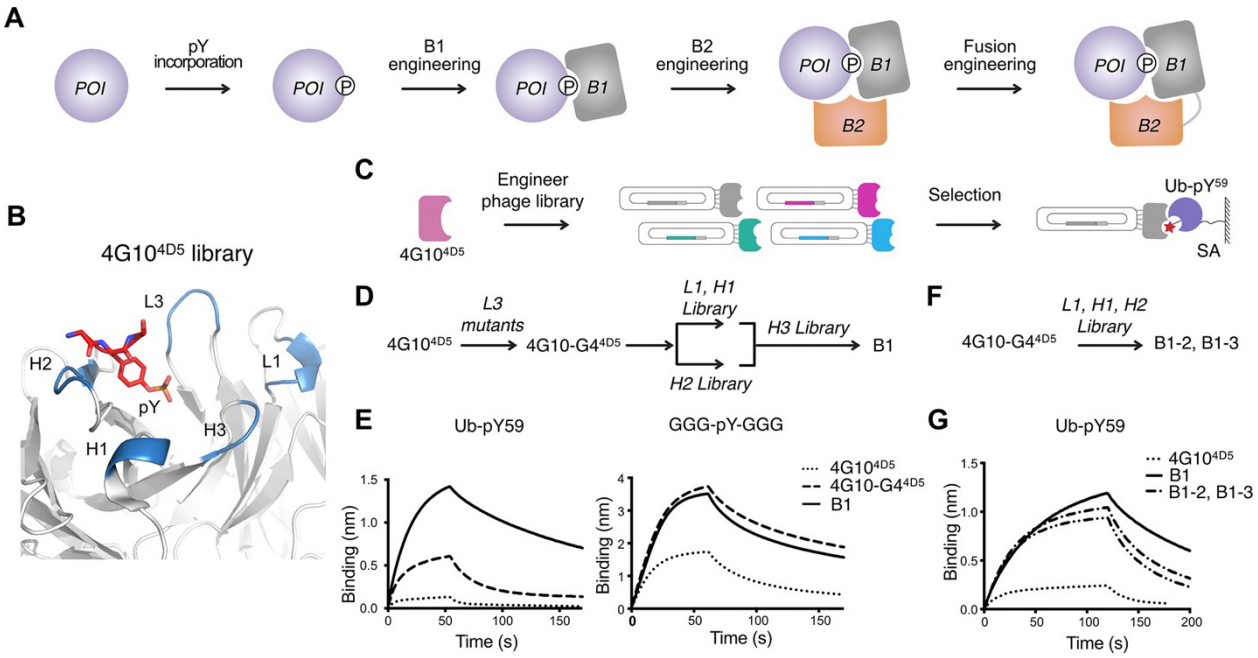


Figure 1.2: Anti-Ub-pY59 B1 engineering. (A) Four-step process to create pY-TRAPs to pY-native proteins: first, generate a protein of interest (POI) with the single pY modification; second, identify B1 that interacts with the pY motif and the surrounding sequence; third, identify B2 that conditionally binds to the complex of B1 and the POI at a composite site which includes both the POI and B1, and not either alone; and finally, engineer B1–B2 fusions to further improve binding affinity. (B) Structure of the 4G10^{4D5} scaffold (PDB: 6DF1) showing the CDR loops (blue) and the pY residue (red). (C) Phage display workflow for anti-Ub-pY59 B1 selection. (D) CDR-walking approach to engineer anti-Ub-pY59 B1. (E) BLI characterization of 4G10^{4D5}, 4G10-G4^{4D5}, and B1 to Ub-pY59 and GGG-pY-GGG. (F) Single library approach where CDRs L1, H1, and H2 are simultaneously mutated to engineer anti-Ub-pY59 B1 binders. (G) BLI characterization of 4G10^{4D5}, B1, B1–2, and B1–3. The K_Ds of the interactions are summarized in Table S1.

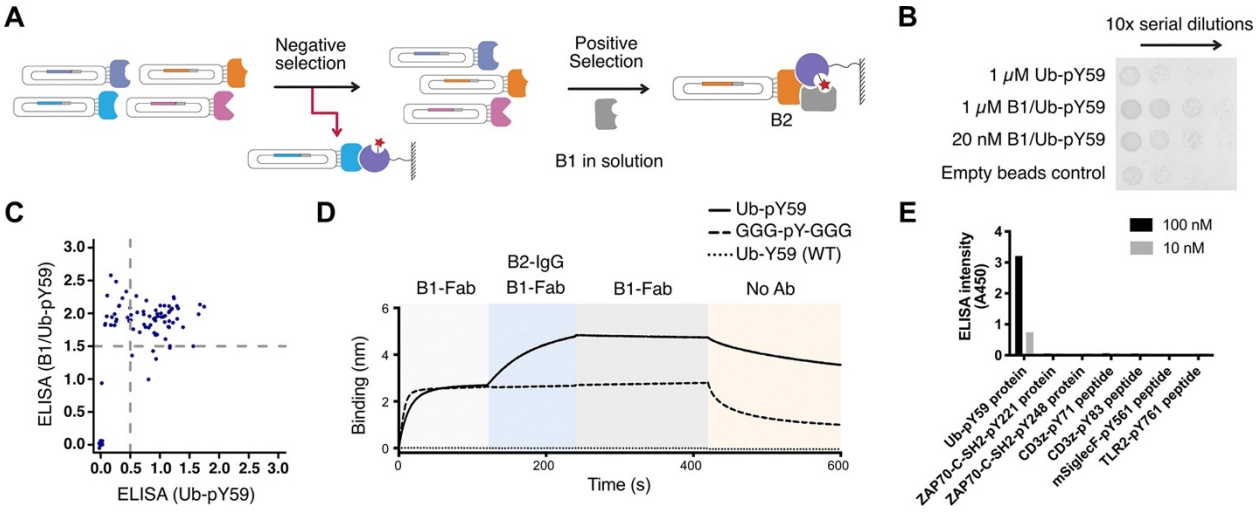
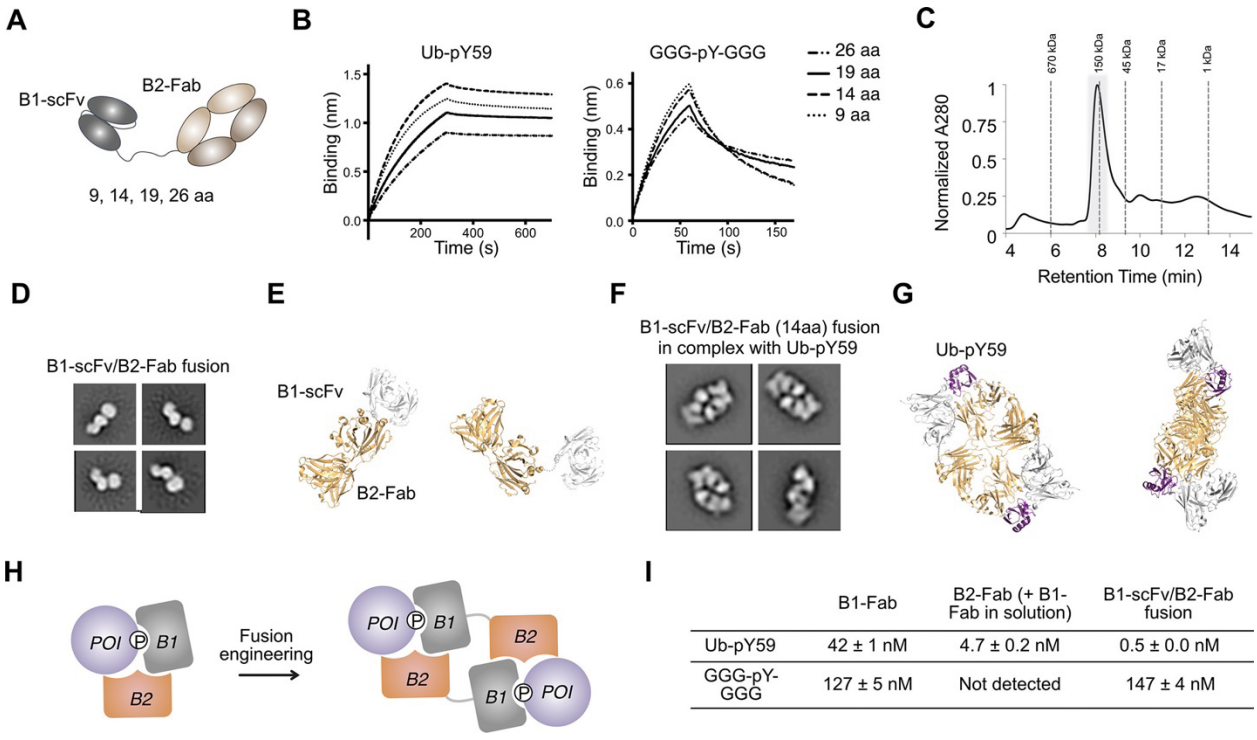
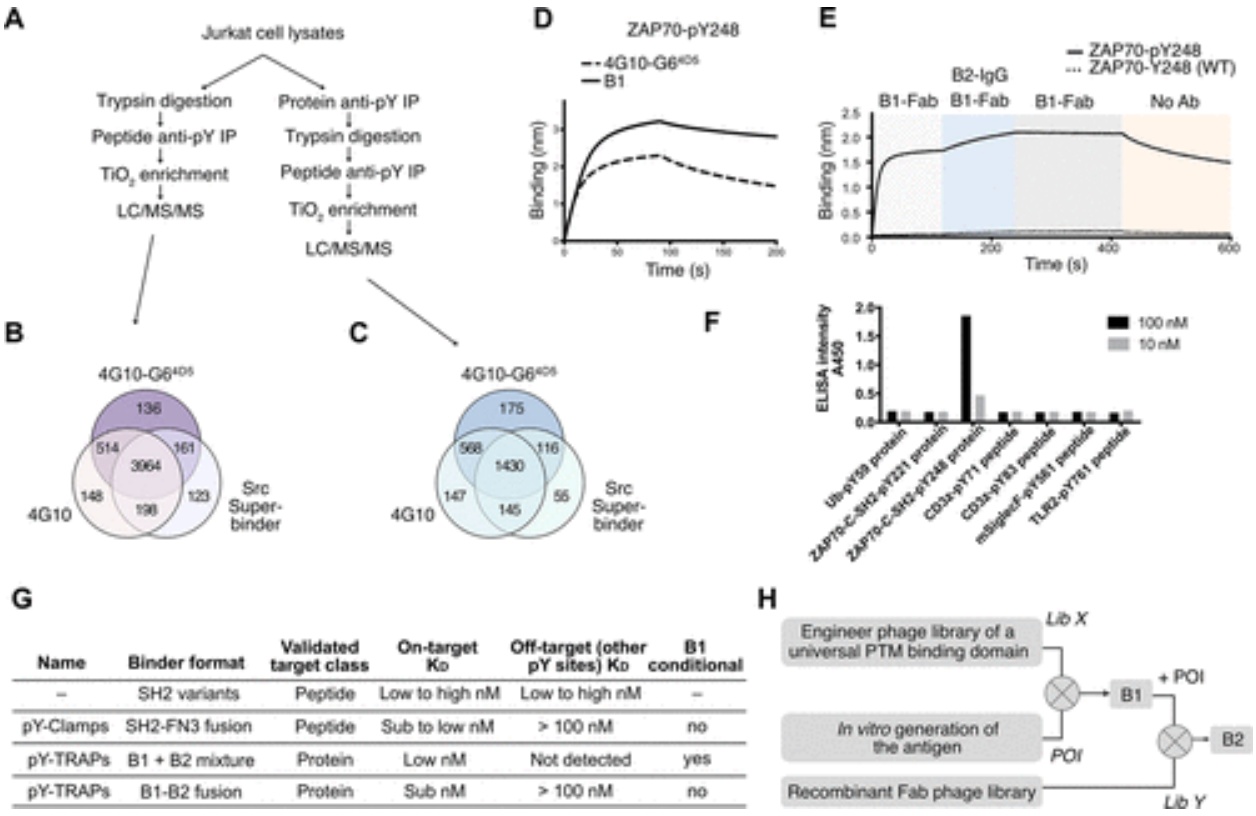


Figure 1.3: Anti-Ub-pY59 B2 engineering dramatically enhances affinity and specificity. (A) Phage display workflow for negative and positive selection for anti-Ub-pY59 B2. (B) Enrichment phage binders as a function of round by phage titer experiments. Binders that strongly interact with the B1/Ub-pY59 complex were more enriched than binders against Ub-pY59 as seen by higher numbers of phagemid colonies. (C) Characterization of binding of Fab-phage to Ub-pY59 or B1/Ub-pY59 in ELISA revealed clones that interact with both the Ub-pY59 and the B1/Ub-pY59 complex (upper right quadrant), and clones that selectively bind to the B1/Ub-pY59 complex (upper left quadrant). (D) Sequential BLI experiments show that B1 binds both Ub-pY59 (solid curve) and the GGG-pY-GGG control (dashed curve), while B2 added subsequently only recognizes the B1/Ub-pY59 complex (solid curve) but not the B1/GGG-pY-GGG complex (dashed curve); neither B1 nor B2 binds to the WT Ub-Y59 protein (dotted curve). (E) ELISA experiment showing that the anti-Ub-pY59 TRAP binders are highly selective toward Ub-pY59 and not other pY proteins or peptides such as the human ZAP70 SH-2 domain modified at two sites, CD3ζ modified at two sites, and TLR2 and murine SiglecF (mSiglecF) modified at one site.



	B1-Fab	B2-Fab (+ B1-Fab in solution)	B1-scFv/B2-Fab fusion
Ub-pY59	42 ± 1 nM	4.7 ± 0.2 nM	0.5 ± 0.0 nM
GGG-pY-GGG	127 ± 5 nM	Not detected	147 ± 4 nM

Figure 1.4: B1-scFv/B2-Fab fusion engineering. (A) Cartoon illustration of the B1-scFv/B2-Fab fusion. Various linker lengths (9, 14, 19, and 26 residues) were tested in the fusion protein between B1 and B2. (B) BLI experiments show that B1-scFv/B2-Fab with a 14-aa linker has the biggest difference in binding affinity for Ub-pY59 and GGG-pY-GGG. (C) The SEC analysis of the complex of B1-scFv/B2-Fab and Ub-pY59 shows that the complex has a molecular weight ~150 kDa. The peak (highlighted in gray background) is collected for SDS-PAGE and NS-EM analysis. (D) Representative 2D class averages of NS-EM data for the B1-scFv/B2-Fab fusion alone. (E) Structural models of the B1-scFv/B2-Fab fusion. Light orange, Fab, PDB 1N8Z; gray, scFv, PDB 6DF1. (F) Representative 2D class averages of NS-EM data for B1-scFv/B2-Fab in complex with Ub-pY59. (G) Structure models of a 2:2 dimer of B1-scFv/B2-Fab and Ub-pY59. B1-scFv and B2-Fab in one polypeptide chain interact with two different Ub-pY59 molecules. Light orange, Fab, PDB 1N8Z; gray, scFv, PDB 6DF1; purple, Ub, PDB 5XK5. (H) Cartoon illustration of the formation of a 2:2 dimer of B1-scFv/B2-Fab and Ub-pY59. (I) Summary of K_D values for B1-Fab, B2-Fab, and B1-scFv/B2-Fab to Ub-pY59 and the GGG-pY-GGG peptide.



Name	Binder format	Validated target class	On-target K_D	Off-target (other pY sites) K_D	B1 conditional
–	SH2 variants	Peptide	Low to high nM	Low to high nM	–
pY-Clamps	SH2-FN3 fusion	Peptide	Sub to low nM	> 100 nM	no
pY-TRAPs	B1 + B2 mixture	Protein	Low nM	Not detected	yes
pY-TRAPs	B1-B2 fusion	Protein	Sub nM	> 100 nM	no

Figure 1.5: The TRAP platform has the potential to generate binders to thousands of other pY sequences and different protein PTMs. (A) Phosphoproteomics workflow to globally identify pY sequences from Jurkat cell lysates by pY peptide immunoprecipitation (IP) or pY protein+peptide IP. (B) Number of pY sequences identified in LC/MS/MS following peptide IP and TiO_2 enrichment by 4G10, 4G10-G6^{4D5}, or Src-Superbinder SH2 domain. (C) Number of pY sequences identified in LC/MS/MS following protein IP, peptide IP, and TiO_2 enrichment by 4G10, 4G10-G6^{4D5}, or Src-Superbinder SH2 domain. (D) The BLI characterization shows that ZAP70-pY248 binds tighter with the identified B1 than with parental 4G10-G6^{4D5}. (E) Sequential BLI experiments show that B1 binds ZAP70-pY248 (solid curve) but not ZAP70-Y248 (WT) (dotted curve), and B2-IgG added subsequently recognizes the B1/ZAP70-pY248 complex (solid curve) but not the ZAP70-Y248 (WT) (dotted curve). (F) The ELISA experiment showed that the anti-ZAP70-pY248 TRAP binders are highly selective against ZAP70-pY248 and not other pY proteins/peptides. (G) Comparison of methods for engineering sequence-specific pY binders. pY-TRAPs can be used in two formats. The B1 plus B2 format provided the highest selectivity while the B1–B2 fusion format resulted in a tighter binder. Except for pY-TRAPs, the other methods generated binders against the linear form of a pY antigen. (H) Proposed workflow to engineer TRAP binders against other protein PTMs.

1.7 Extended and Supplemental Figures

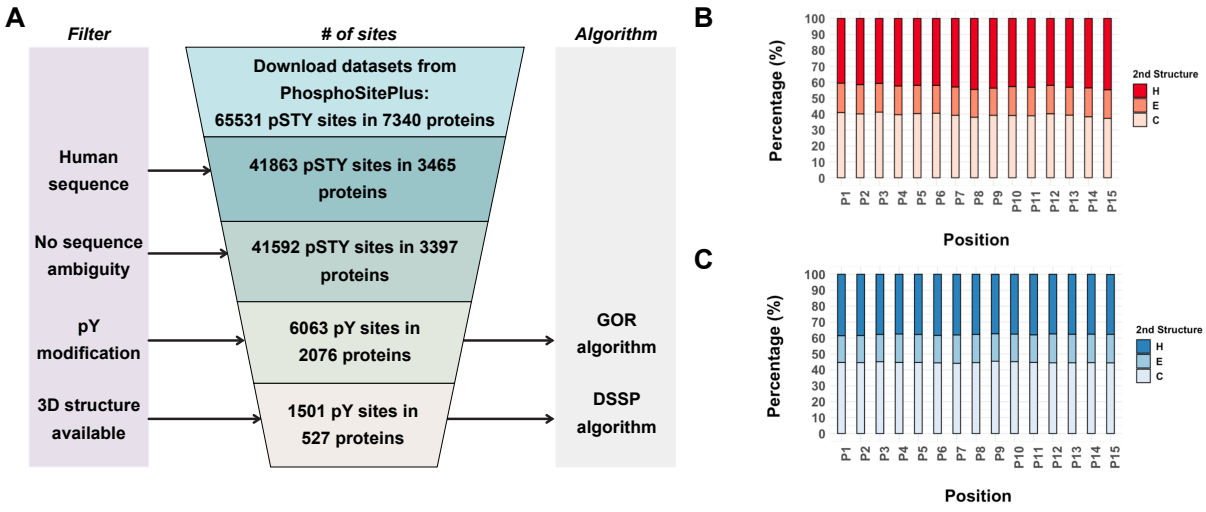


Figure 1.S1: Bioinformatics analysis of secondary structures for tyrosine kinase substrate sequences suggests high proportion of pY sites are in classic regions of secondary structure. (A) The process of secondary structure analysis for pY sequences banked at PhosphoSitePlus. In total, 1501 human pY sequences in 527 proteins which have available parental protein PDB structures were analyzed using the DSSP algorithm, and 6063 human pY sequences in 2076 proteins were analyzed using the GOR algorithm. (B) Secondary structure prediction of randomly selected 15-residue sequences from proteins analyzed in Figure 1B using the DSSP algorithm. (C) Secondary structure analysis of randomly selected 15-residue sequences from proteins analyzed in Figure 1C using the GOR algorithm.

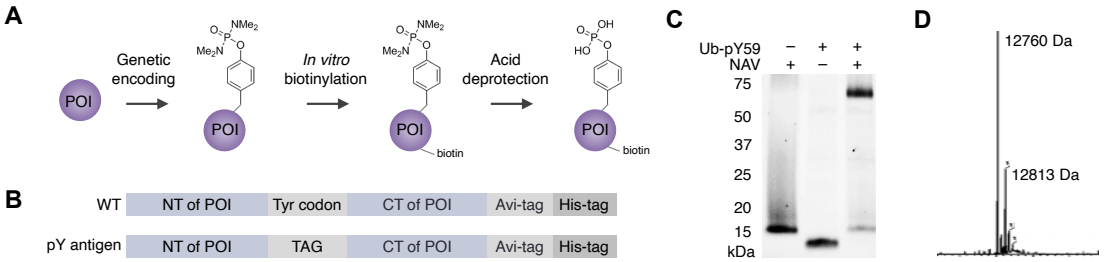


Figure 1.S2: Generation of Ub-pY59 used for phage selections. (A) The workflow to generate biotinylated Ub-pY59 using an unnatural amino acid method. (B) Constructs for expressing Ub-Y59 (WT) and Ub-pY59. (C) SDS-PAGE analysis of Ub-pY59 in the absence (lane 2) or presence (lane 3) of neutravidin (NAV) confirmed Ub-pY59 was fully biotinylated. (D) LC-MS validation of the biotinylated Ub-pY59 protein. The Ub protein with the protected pY Uaa incorporated at site 59: 12813 Da; the deprotected Ub-pY59: 12760 Da.

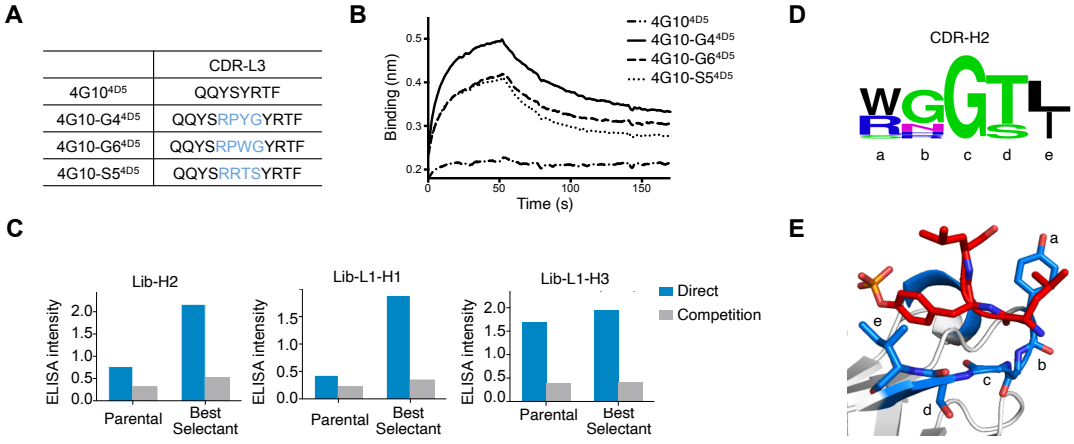


Figure 1.S3: Engineering of anti-Ub-pY59 B1. (A) The CDR L3 sequences in the 4G104D5 selectants. (B) BLI experiments show 4G10-G44D5 binds the tightest to Ub-pY59 among the 4G104D5 variants tested. (C) Direct and competition ELISA intensities comparing parental clone and the best variants identified from each library. (D)(E) Amino acid sequences are highly conserved for site c, d, and e in CDR H2.

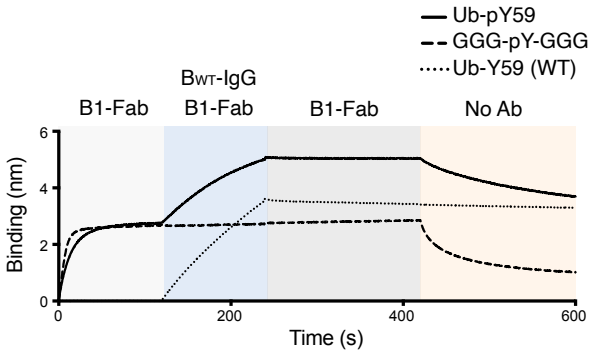


Figure 1.S4: Sequential BLI experiments shows a non-conditional binder BWT that interacts with both the Ub-Y59 (WT) and Ub-pY59 in complex with B1. The solid curve shows binding of B1-Fab to Ub-pY59, followed by binding of BWT-IgG. The dashed curve shows binding of the B1-Fab to the immobilized GGG-pY-GGG, but no binding of BWT-IgG when added subsequently. The dotted curve shows no binding of B1-Fab to Ub-Y59 (WT) but binding of BWT to Ub-Y59 (WT) in the subsequent step.

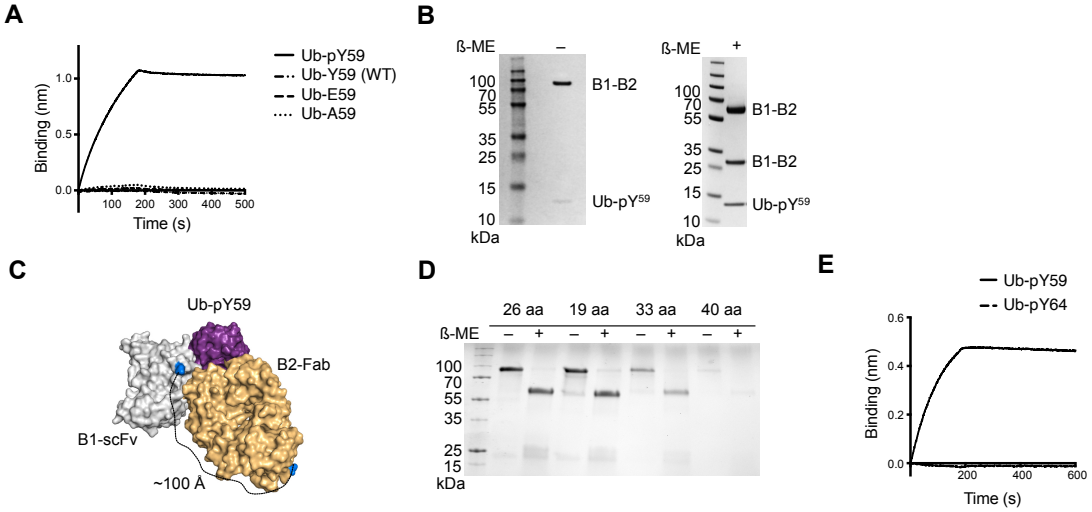


Figure 1.S5: Characterization of B1-scFv/B2-Fab fusions. (A) B1-scFv/B2-Fab (14aa) does not interact with Ub-Y59 (WT), Ub-A59, Ub-E59. (B) SDS-PAGE gel analysis of the peak fraction in collected from SEC experiment described in Figure 4C. (C) Structure models of a 1:1 dimer of B1-scFv/B2-Fab and Ub-pY59. A long-linker is required for B1-scFv and B2-Fab in one polypeptide chain to interact with a single Ub-pY59 molecule. Light orange: Fab, PDB 1N8Z; grey: scFv, PDB 6DF1; purple: Ub, PDB 5XK5. (D) SDS-Page analysis of the same volume of purified B1/B2 fusion proteins containing various linker lengths showed that the yield of proteins decreased significantly with longer linkers. (F) B1-scFv/B2-Fab (14aa) does not interact with Ub-pY64.

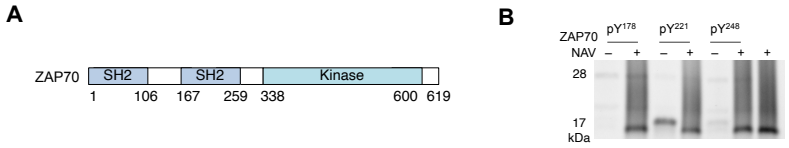


Figure 1.S6: Generation of ZAP70-C-SH2 (aa 167-259) pY antigens. (A) Schematics of the ZAP70 protein domains. (B) SDS-PAGE gel shows expression of ZAP70-C-SH2 (aa 167-259)-pY221 and pY248 at the correct molecular weight (~17 kDa). Gel analysis in the absence (lane 3 and 5) or presence (lane 4 and 6) of NAV confirmed ZAP70-pY221 and 248 were fully biotinylated.

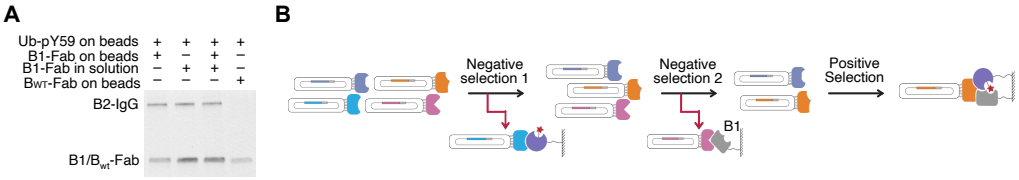


Figure 1.S7: Optimization of the B2 engineering workflow. (A) Immunoprecipitation experiment shows immobilizing both B1-Fab and pY protein on SA beads generated a stable B1/pY antigen complex for interacting with B2-IgG. Adding extra B1-Fab in solution did not pull down more B2-IgG. (B) Diagram of the optimized phage selection strategy for B2 engineering. A recombinant Fab-phage library is first cleared with immobilized pY-protein (purple), and then cleared with immobilized B1 (grey), followed by selection against B1/pY-antigen complex with both B1 and pY-antigen co-immobilized on the SA beads to pull down complex-specific binders.

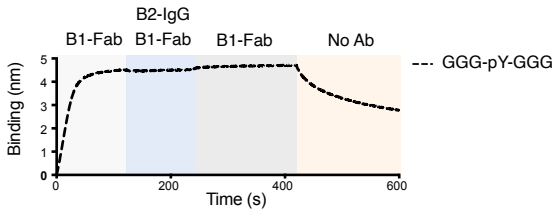


Figure 1.S8: Sequential BLI characterization of anti-ZAP70-pY248 binders on GGG-pY-GGG control peptide shows B2 does not bind reflecting its selectivity. The dashed curve shows binding of the B1-Fab to the immobilized GGG-pY-GGG, but no binding of B2-IgG when added subsequently.

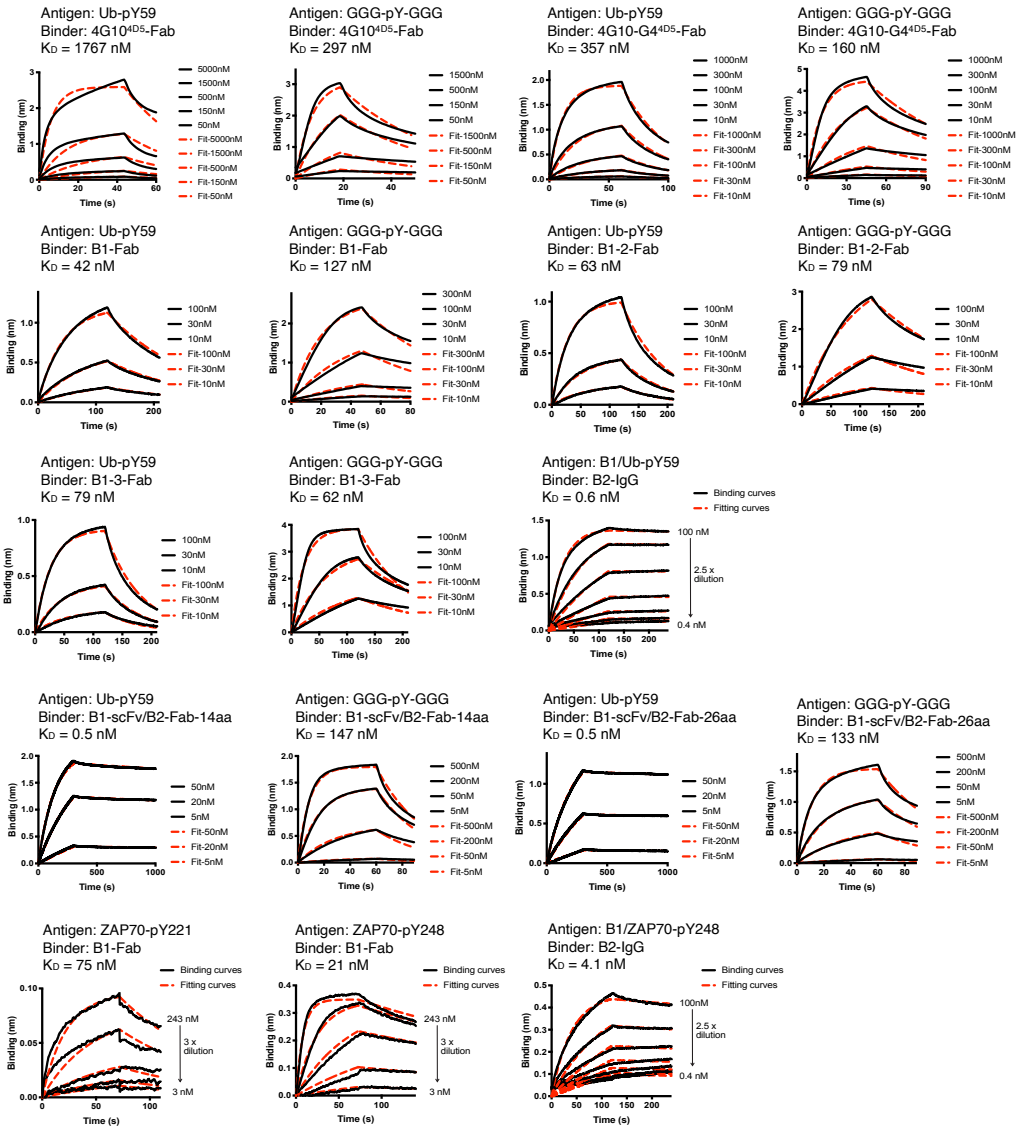


Figure 1.S9: BLI measurements for the pY-TRAP antibodies binding to the pY-antigens. Dotted lines show calculated fits. The KD values and the error of fits are summarized in Table S1.

Table 1.S1: Summary of KDs determined by BLI measurements [a]

Binder	Antigen 1	K_D (nM) for Antigen 1	Antigen 2	K_D (nM) for Antigen 2
4G10^{4D5}	Ub-pY59	1767± 139	GGG-pY-GGG	297 ± 13
4G10-G4^{4D5}	Ub-pY59	357 ± 7	GGG-pY-GGG	160 ± 5
Anti-Ub-pY59 B1-Fab	Ub-pY59	42 ± 1	GGG-pY-GGG	127 ± 5
Anti-Ub-pY59 B1-2-Fab	Ub-pY59	63 ±1	GGG-pY-GGG	79 ± 4
Anti-Ub-pY59 B1-3-Fab	Ub-pY59	79 ± 2	GGG-pY-GGG	62 ± 3
Anti-Ub-pY59 B2-Fab	B1/Ub-pY59	4.7 ± 0.2[b]	B1/GGG-pY-GGG	Undetectable
Anti-Ub-pY59 B2-IgG	B1/Ub-pY59	0.6 ± 0.1	B1/GGG-pY-GGG	Undetectable
Anti-Ub-pY59 B1-scFv/B2-Fab-14aa	Ub-pY59	0.5 ± 0.0	GGG-pY-GGG	147 ± 4
Anti-Ub-pY59 B1-scFv/B2-Fab-26aa	Ub-pY59	0.5 ± 0.0	GGG-pY-GGG	133 ± 3
4G10-G6^{4D5}	ZAP70-pY221	73 ± 4[b]	GGG-pY-GGG	Not determined
Anti-ZAP70-pY221 B1-Fab	ZAP70-pY221	75 ± 4	GGG-pY-GGG	Not determined
4G10-G6^{4D5}	ZAP70-pY248	56 ± 1[b]	GGG-pY-GGG	Not determined
Anti-ZAP70-pY248 B1-Fab	ZAP70-pY248	21 ± 2	GGG-pY-GGG	Not determined
Anti-ZAP70-pY248 B2-IgG	B1/ZAP70-pY248	4.1 ± 0.1	GGG-pY-GGG	Undetectable

[a]Table S1 lists the K_D values with errors of the fits for titration curves shown in Figure S9. All K_D values were determined by multi-concentration BLI measurements excepted for those noted.

[b]These K_Ds were estimated based on single-concentration BLI measurement.

Table 1.S 2. Sequences of CDRs [a,b]

Name	CDR L1	CDR L2	CDR L3	CDR H1	CDR H2	CDR H3
4G10^{4D5}	RASSSVSSS YLH	STSNLAS	YSYR	(YT)FTENTV	(IG)GINPYYGGSI(FSPKF)	RAGAY(YF)
4G10-G4^{4D5}	RASSSVSSS YLH	STSNLAS	YSRPYGYR	(YT)FTENTV	(IG)GINPYYGGSI (FSPKF)	RAGAY(YF)
4G10-G6^{4D5}	RASSSVSSS YLH	STSNLAS	YSRPWGYR	(YT)FTENTV	(IG)GINPYYGGSI(FSPKF)	RAGAY(YF)
4G10-S5^{4D5}	RASSSVSSS YLH	STSNLAS	YSRRTSYR	(YT)FTENTV	(IG)GINPYYGGSI(FSPKF)	RAGAY(YF)
Anti-Ub-pY59 B1	RASSSVFPS VLH	STSNLAS	YSRPYGYR	(YT)FDRHTV	(IG)GINPYWGGTI(FSPKF)	RAGAY(YF)
Anti-Ub-pY59 B2	RASQSVGSA LA	SASSLYS	ASGYLF	LSYYYI	SIYPSYSSTY	AYYFGSGL
Anti-Ub B_{WT}	RASQSVGSA LA	SASSLYS	SRAYSLF	IYSYSM	YISPYYGSTS	YFSWGWDA F
Anti-ZAP70-pY248 B1	RASSSV RFKWLH	STSNLAS	YSRPWGYR	(YT)FINNTV	(IG)GINPYYGGSI(FSPKF)	RAGAY(YF)
Anti-ZAP70-pY248 B2	RASQSVGSA LA	SASSLYS	YYHRPF	LSYSSI	SISSYYGSTS	PFYGSYFYA I

[a]Amino acids denoted in brackets of CDR H1, H2, and H3 are flanking constant regions that differ from canonical "Herceptin" 4D5 scaffold. This was a design consideration during CDR grafting from 4G10 to "Herceptin" 4D5 scaffold detailed in work by Mou *et al*.[5]

[b]Amino acids in blue: sites randomized in phage library described in Table S3.

Table 1.S3. Library designs [a]

Name	Parental	CDR L1	CDR L2	CDR L3	CDR H1	CDR H2	CDR H3	Theoretical diversity	Actual diversity
Lib-H2	4G10-G4^{4D5}	–	–	–	–	NNK VVC NNK VVC NNK	–	$3x10^6$	2×10^7
Lib-L1-H1	4G10-G4^{4D5}	KHT VVC KHT NNK	–	–	KHT NNK NNK	–	–	$6.4x10^7$	3×10^9
Lib-H3	4G10-G4^{4D5}	–	–	–	–	–	$(NNK)_{3-7}$	$3.5x10^{10}$	5×10^8
Lib-L1-H1-H2-1	4G10-G4^{4D5}	NNK NNK NNK NNK	–	–	NNK NNK NNK	NNK NNK NNK NNK NNK	–	1×10^{18}	5×10^9
Lib-L1-H1-H2-2	4G10-G6^{4D5}	NNK NNK NNK NNK	–	–	NNK NNK NNK	NNK NNK	–	3.5×10^{13}	5×10^9

[a]Mixed bases code for DNA: N (A,C,G,T); K (G,T); V (A,C,G); K (G,T); H (A,C,T).

1.8 References

1. Hunter, T. Tyrosine phosphorylation: thirty years and counting. *Curr. Opin. Cell Biol.* **2009**, *21*, 140– 146, DOI: 10.1016/j.ceb.2009.01.028
2. Lim, W. A.; Pawson, T. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell* **2010**, *142*, 661– 667, DOI: 10.1016/j.cell.2010.08.023
3. Hornbeck, P. V.; Kornhauser, J. M.; Tkachev, S.; Zhang, B.; Skrzypek, E.; Murray, B.; Latham, V.; Sullivan, M. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* **2012**, *40*, D261– 70, DOI: 10.1093/nar/gkr1122
4. Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **2015**, *43*, D512– D520, DOI: 10.1093/nar/gku1267
5. Mou, Y.; Zhou, X. X.; Leung, K.; Martinko, A. J.; Yu, J.-Y.; Chen, W.; Wells, J. A. Engineering Improved Antiphosphotyrosine Antibodies Based on an Immunoconvergent Binding Motif. *J. Am. Chem. Soc.* **2018**, *140*, 16615– 16624, DOI: 10.1021/jacs.8b08402
6. Hattori, T.; Koide, S. Next-generation antibodies for post-translational modifications. *Curr. Opin. Struct. Biol.* **2018**, *51*, 141– 148, DOI: 10.1016/j.sbi.2018.04.006
7. Koerber, J. T.; Thomsen, N. D.; Hannigan, B. T.; Degrado, W. F.; Wells, J. A. Nature-inspired design of motif-specific antibody scaffolds. *Nat. Biotechnol.* **2013**, *31*, 916– 921, DOI: 10.1038/nbt.2672
8. Brumbaugh, K.; Johnson, W.; Liao, W. C.; Lin, M. S.; Houchins, J. P.; Cooper, J.; Stoesz, S.; Campos-Gonzalez, R. Overview of the generation, validation, and application of

phosphosite-specific antibodies. *Methods Mol. Biol.* **2011**, *717*, 3– 43, DOI: 10.1007/978-1-61779-024-9_1

9. Kehoe, J. W.; Velappan, N.; Walbolt, M.; Rasmussen, J.; King, D.; Lou, J.; Knopp, K.; Pavlik, P.; Marks, J. D.; Bertozzi, C. R.; Bradbury, A. R. Using phage display to select antibodies recognizing post-translational modifications independently of sequence context. *Mol. Cell. Proteomics* **2006**, *5*, 2350– 2363, DOI: 10.1074/mcp.M600314-MCP200
10. Mann, J. K.; Wood, J. F.; Stephan, A. F.; Tzanakakis, E. S.; Ferkey, D. M.; Park, S. Epitope-guided engineering of monobody binders for in vivo inhibition of Erk-2 signaling. *ACS Chem. Biol.* **2013**, *8*, 608– 616, DOI: 10.1021/cb300579e
11. Kaneko, T.; Huang, H.; Cao, X.; Li, X.; Li, C.; Voss, C.; Sidhu, S. S.; Li, S. S. Superbinder SH2 domains act as antagonists of cell signaling. *Sci. Signaling* **2012**, *5*, ra68 DOI: 10.1126/scisignal.2003021
12. Malabarba, M. G.; Milia, E.; Faretta, M.; Zamponi, R.; Pelicci, P. G.; Di Fiore, P. P. A repertoire library that allows the selection of synthetic SH2s with altered binding specificities. *Oncogene* **2001**, *20*, 5186– 5194, DOI: 10.1038/sj.onc.1204654
13. Yasui, N.; Findlay, G. M.; Gish, G. D.; Hsiung, M. S.; Huang, J.; Tucholska, M.; Taylor, L.; Smith, L.; Boldridge, W. C.; Koide, A.; Pawson, T.; Koide, S. Directed network wiring identifies a key protein interaction in embryonic stem cell differentiation. *Mol. Cell* **2014**, *54*, 1034– 1041, DOI: 10.1016/j.molcel.2014.05.002
14. Durek, P.; Schudoma, C.; Weckwerth, W.; Selbig, J.; Walther, D. Detection and characterization of 3D-signature phosphorylation site motifs and their contribution towards

improved phosphorylation site prediction in proteins. *BMC Bioinf.* **2009**, *10*, 117, DOI: 10.1186/1471-2105-10-117

15. Jiménez, J. L.; Hegemann, B.; Hutchins, J. R.; Peters, J. M.; Durbin, R. A systematic comparative and structural analysis of protein phosphorylation sites based on the mtcPTM database. *Genome Biol.* **2007**, *8*, R90, DOI: 10.1186/gb-2007-8-5-r90
16. Hoppmann, C.; Wong, A.; Yang, B.; Li, S.; Hunter, T.; Shokat, K. M.; Wang, L. Site-specific incorporation of phosphotyrosine using an expanded genetic code. *Nat. Chem. Biol.* **2017**, *13*, 842– 844, DOI: 10.1038/nchembio.2406
17. Consortium, U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506– D515, DOI: 10.1093/nar/gky1049
18. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577– 2637, DOI: 10.1002/bip.360221211
19. Garnier, J.; Gibrat, J.-F.; Robson, B. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* **1996**, *266*, 540– 553, DOI: 10.1016/S0076-6879(96)66034-0
20. Swatek, K. N.; Komander, D. Ubiquitin modifications. *Cell Res.* **2016**, *26*, 399– 422, DOI: 10.1038/cr.2016.39
21. Yau, R.; Rape, M. The increasing complexity of the ubiquitin code. *Nat. Cell Biol.* **2016**, *18*, 579– 586, DOI: 10.1038/ncb3358
22. Chong, R. A.; Wu, K.; Spratt, D. E.; Yang, Y.; Lee, C.; Nayak, J.; Xu, M.; Elkholi, R.; Tappin, I.; Li, J.; Hurwitz, J.; Brown, B. D.; Chipuk, J. E.; Chen, Z. J.; Sanchez, R.; Shaw, G. S.; Huang, L.; Pan, Z. Q. Pivotal role for the ubiquitin Y59-E51 loop in lysine

48 polyubiquitination. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8434– 8439, DOI: 10.1073/pnas.1407849111

23. Luo, X.; Fu, G.; Wang, R. E.; Zhu, X.; Zambaldo, C.; Liu, R.; Liu, T.; Lyu, X.; Du, J.; Xuan, W.; Yao, A.; Reed, S. A.; Kang, M.; Zhang, Y.; Guo, H.; Huang, C.; Yang, P. Y.; Wilson, I. A.; Schultz, P. G.; Wang, F. Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* **2017**, *13*, 845– 849, DOI: 10.1038/nchembio.2405

24. Fairhead, M.; Howarth, M. Site-specific biotinylation of purified proteins using BirA. *Methods Mol. Biol.* **2015**, *1266*, 171– 184, DOI: 10.1007/978-1-4939-2272-7_12

25. Eisinger, D.; Stiles, L.; LaMarche, A.; Jelinek, T. *Recombinant monoclonal antibody to phosphotyrosine-containing proteins*. US patent US 6824989 B1, September 1, 2000.

26. Yang, W.-P.; Green, K.; Pinz-Sweeney, S.; Briones, A. T.; Burton, D. R.; Barbas III, C. F. CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range. *J. Mol. Biol.* **1995**, *254*, 392– 403, DOI: 10.1006/jmbi.1995.0626

27. Almagro, J. C. Identification of differences in the specificity-determining residues of antibodies that recognize antigens of different size: implications for the rational design of antibody repertoires. *J. Mol. Recognit.* **2004**, *17*, 132– 143, DOI: 10.1002/jmr.659

28. Xu, J. L.; Davis, M. M. Diversity in the CDR3 region of VH is sufficient for most antibody specificities. *Immunity* **2000**, *13*, 37– 45, DOI: 10.1016/S1074-7613(00)00006-6

29. Persson, H.; Ye, W.; Wernimont, A.; Adams, J. J.; Koide, A.; Koide, S.; Lam, R.; Sidhu, S. S. CDR-H3 diversity is not required for antigen recognition by synthetic antibodies. *J. Mol. Biol.* **2013**, *425*, 803– 811, DOI: 10.1016/j.jmb.2012.11.037

30. George, R. A.; Heringa, J. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng., Des. Sel.* **2002**, *15*, 871– 879, DOI: 10.1093/protein/15.11.871

31. Bian, Y.; Li, L.; Dong, M.; Liu, X.; Kaneko, T.; Cheng, K.; Liu, H.; Voss, C.; Cao, X.; Wang, Y. Ultra-deep tyrosine phosphoproteomics enabled by a phosphotyrosine superbinder. *Nat. Chem. Biol.* **2016**, *12*, 959– 966, DOI: 10.1038/nchembio.2178

32. Szabo, M.; Czompoly, T.; Kvell, K.; Talaber, G.; Bartis, D.; Nemeth, P.; Berki, T.; Boldizsar, F. Fine-tuning of proximal TCR signaling by ZAP-70 tyrosine residues in Jurkat cells. *Int. Immunol.* **2012**, *24*, 79– 87, DOI: 10.1093/intimm/dxr105

33. Helou, Y. A.; Nguyen, V.; Beik, S. P.; Salomon, A. R. ERK positive feedback regulates a widespread network of tyrosine phosphorylation sites across canonical T cell signaling and actin cytoskeletal proteins in Jurkat T cells. *PLoS One* **2013**, *8*, e69641 DOI: 10.1371/journal.pone.0069641

34. Spiess, C.; Zhai, Q.; Carter, P. J. Alternative molecular formats and therapeutic applications for bispecific antibodies. *Mol. Immunol.* **2015**, *67*, 95– 106, DOI: 10.1016/j.molimm.2015.01.003

35. de Oliveira, P. S.; Ferraz, F. A.; Pena, D. A.; Pramio, D. T.; Morais, F. A.; Schechtman, D. Revisiting protein kinase-substrate interactions: Toward therapeutic development. Revisiting protein kinase-substrate interactions: Toward therapeutic development. *Sci. Signaling* **2016**, *9*, re3 DOI: 10.1126/scisignal.aad4016

36. Kaneko, T.; Joshi, R.; Feller, S. M.; Li, S. S. C. Phosphotyrosine recognition domains: the typical, the atypical and the versatile. *Cell Commun. Signaling* **2012**, *10*, 32, DOI: 10.1186/1478-811X-10-32

37. Duarte, M. L.; Pena, D. A.; Nunes Ferraz, F. A.; Berti, D. A.; Paschoal Sobreira, T. J.; Costa-Junior, H. M.; Abdel Baqui, M. M.; Disatnik, M. H.; Xavier-Neto, J.; Lopes de Oliveira, P. S.; Schechtman, D. Protein folding creates structure-based, noncontiguous consensus phosphorylation motifs recognized by kinases. *Sci. Signaling* **2014**, *7*, ra105 DOI: 10.1126/scisignal.2005412

38. Trost, B.; Kusalik, A. Computational prediction of eukaryotic phosphorylation sites. *Bioinformatics* **2011**, *27*, 2927– 2935, DOI: 10.1093/bioinformatics/btr525

39. Hattori, T.; Lai, D.; Dementieva, I. S.; Montaño, S. P.; Kurosawa, K.; Zheng, Y.; Akin, L. R.; Świst-Rosowska, K. M.; Grzybowski, A. T.; Koide, A.; Krajewski, K.; Strahl, B. D.; Kelleher, N. L.; Ruthenburg, A. J.; Koide, S. Antigen clasping by two antigen-binding sites of an exceptionally specific antibody for histone methylation. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2092– 2097, DOI: 10.1073/pnas.1522691113

40. Ju, T.; Niu, W.; Guo, J. Evolution of Src Homology 2 (SH2) Domain to Recognize Sulfotyrosine. *ACS Chem. Biol.* **2016**, *11*, 2551– 2557, DOI: 10.1021/acschembio.6b00555

41. Lawrie, J.; Niu, W.; Guo, J. Engineering of a sulfotyrosine-recognizing small protein scaffold for the study of protein tyrosine O-sulfation. *Methods Enzymol.* **2019**, *622*, 67– 89, DOI: 10.1016/bs.mie.2019.02.004

42. Cummings, R. D.; Etzler, M. E. Antibodies and Lectins in Glycan Analysis. In *Essentials of Glycobiology*; Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2009; Chapter 45.

43. Zhang, L.; Luo, S.; Zhang, B. The use of lectin microarray for assessing glycosylation of therapeutic proteins. *MAbs* **2016**, *8*, 524– 535, DOI: 10.1080/19420862.2016.1149662

44. Takimoto, J. K.; Dellas, N.; Noel, J. P.; Wang, L. Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids. *ACS Chem. Biol.* **2011**, *6*, 733– 743, DOI: 10.1021/cb200057a

45. Kunkel, T. A.; Bebenek, K.; McClary, J. Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* **1991**, *204*, 125– 139, DOI: 10.1016/0076-6879(91)04008-C

46. Hornsby, M.; Paduch, M.; Miersch, S.; Sääf, A.; Matsuguchi, T.; Lee, B.; Wypisniak, K.; Doak, A.; King, D.; Usatyuk, S.; Perry, K.; Lu, V.; Thomas, W.; Luke, J.; Goodman, J.; Hoey, R. J.; Lai, D.; Griffin, C.; Li, Z.; Vizeacoumar, F. J.; Dong, D.; Campbell, E.; Anderson, S.; Zhong, N.; Gräslund, S.; Koide, S.; Moffat, J.; Sidhu, S.; Kossiakoff, A.; Wells, J. A High Through-put Platform for Recombinant Antibodies to Folded Proteins. *Mol. Cell. Proteomics* **2015**, *14*, 2833– 2847, DOI: 10.1074/mcp.O115.052209

47. Ohi, M.; Li, Y.; Cheng, Y.; Walz, T. Negative staining and image classification—powerful tools in modern electron microscopy. *Biol. Proced. Online* **2004**, *6*, 23– 34, DOI: 10.1251/bpo70

48. Booth, D. S.; Avila-Sakar, A.; Cheng, Y. Visualizing proteins and macromolecular complexes by negative stain EM: from grid preparation to image acquisition. *J. Visualized Exp.* **2011**, e3227 DOI: 10.3791/3227

49. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **2016**, *193*, 1– 12, DOI: 10.1016/j.jsb.2015.11.003

50. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **2012**, *180*, 519– 530, DOI: 10.1016/j.jsb.2012.09.006

Chapter 2

Development of Bi-paratopic and multivalent VH domains to block ACE2 binding and neutralize SARS-CoV-2

2.1 Abstract

Neutralizing agents against SARS-CoV-2 are urgently needed for the treatment and prophylaxis of COVID-19. Here, we present a strategy to rapidly identify and assemble synthetic human variable heavy (VH) domains toward neutralizing epitopes. We constructed a VH-phage library and targeted the angiotensin-converting enzyme 2 (ACE2) binding interface of the SARS-CoV-2 Spike receptor-binding domain (Spike-RBD). Using a masked selection approach, we identified VH binders to two non-overlapping epitopes and further assembled these into multivalent and bi-paratopic formats. These VH constructs showed increased affinity to Spike (up to 600-fold) and neutralization potency (up to 1400-fold) on pseudotyped SARS-CoV-2 virus when compared to standalone VH domains. The most potent binder, a trivalent VH, neutralized authentic SARS-CoV-2 with half-maximal inhibitory concentration (IC_{50}) of 4.0 nM (180 ng/mL). A cryo-EM structure of the trivalent VH bound to Spike shows each VH domain engaging an RBD at the ACE2 binding site, confirming our original design strategy.

2.2 Introduction

The emergence of SARS-CoV-2 and the associated COVID-19 disease has emphasized the need to rapidly generate therapeutics against novel pathogens. SARS-CoV-2 enters cells through

the interaction of the viral Spike receptor-binding domain (Spike-RBD) and host angiotensin-converting enzyme-2 (ACE2) on the surface of lung epithelial cells.[1] Antibody and antibody-like biologics that can block this interaction are promising therapeutic candidates because of their high specificity and neutralization potency.[2] The majority of antibodies isolated so far against SARS-CoV-2, SARS-CoV-1, and MERS are derived from screening B-cells of infected patients or repurposed from animal immunizations.[3–7] These approaches, though effective, can be time-consuming and may not necessarily yield neutralizing antibodies. Given the pressing nature of this pandemic, there is a need for multiple parallel strategies to rapidly produce potent, recombinant, and neutralizing biologics.

In vitro display technologies using yeast or phage are well-established approaches for generating high-affinity binders from large naïve libraries.[8] *In vitro* selection can be done without infected individuals and only requires the recombinant protein target. One of the recently developed modalities are small single domain antibodies derived from variable heavy homodimer (VHH) domains of antibodies from camels or llamas, often referred to as nanobodies, and are usually obtained by camelid immunization and B-cell cloning.[9–12] Nanobodies have some advantages. Their single-chain and small size (11 to 15 kDa) allow them to bind epitopes or penetrate tissues that may not be accessible to monoclonal antibodies (mAbs) (150 kDa), and nanobodies can be rapidly produced in *E. coli*.[13,14] However, nanobodies derived from animal immunization can also suffer from long-turnaround times. Although this can be overcome with synthetic nanobody libraries,[15,16] nanobody scaffolds that are animal-derived raise significant concerns regarding immunogenicity. More recently, variable heavy (VH) domains derived from human scaffolds have been produced and tested against a number of targets.[17–19]

Thus, we and others have been interested in developing VH binders to SARS-CoV-2 for the present pandemic, and as a test case for future ones.[20–23] However, one limitation of synthetic single domain binders is that as monomers, they often lack the strong binding affinity necessary for therapeutic application. Affinity maturation can improve this, although this extends the development timeline. Instead, generating linked multivalent or multi-paratopic VH binders could be a more rapid approach to utilize avidity to boost affinity and efficacy.[24] Linking VH domains into such homo- and hetero-bifunctional formats is more straightforward than preparing similar multifunctional antibodies because the latter requires correct heavy and light chain pairing to maintain binding affinity, whereas multifunctional VH domains have no such requirements.[11]

Here, we constructed a human VH-phage library derived from the clinically approved trastuzumab scaffold and validated its use on multiple antigens. By utilizing a masked phage selection strategy, we rapidly identified VH domains at two non-overlapping epitopes within the ACE2 binding site of the SARS-CoV-2 Spike-RBD. By linking these VH domains with a strategic linker into bi-paratopic and multivalent binders, we improved affinity from mid-nM to ~100 pM without any additional high-resolution structural information. These high-affinity binders are capable of potently neutralizing pseudotyped and live SARS-CoV-2. A cryo-electron microscopy (cryo-EM) structure of the most potent trivalent VH bound to Spike shows that each VH domain precisely targets the ACE2 binding interface on all three RBDs of Spike. We believe our VH-phage library and this multivalent and multi-paratopic approach is highly advantageous to target distinct epitopes within an antigen and can be broadly applied to other viral and non-viral targets to leverage avidity for increased potency.

2.3 Results

Construction of a synthetic human VH-phage library

To enable the generation of single-domain antibodies against targets such as SARS-CoV-2 Spike, we designed a synthetic VH-phage library using the VH domain (4D5) from the highly stable and clinically successful trastuzumab antibody (**Fig. 2.1a**).[25,26] The VH scaffold was modified to include five amino acid changes predicted to reduce aggregation (**Supplementary Table 2.1**).[27] To bias toward colloidal stability, aspartate and arginine or glutamate residues were inserted at the beginning and at two or three terminal positions of CDR H1, as these have been previously used to improve aggregation resistance of VH and scFv fragments from the VH3 germline.[19,28] Diversity was introduced into CDR H1 and CDR H2 using a minimalistic approach where variability was largely restricted to tyrosine and serine residues (**Extended Data Fig. 2.1**).[29] We introduced high-diversity mixtures of amino acids into CDR H3 because it is usually critical to antigen recognition (**Extended Data Fig. 2.1**), and Fab-phage libraries with highly diverse CDR H3 sequences have successfully yielded high-affinity antibodies to a variety of target antigens.[30,31] Furthermore, charged polar residues such as aspartate were introduced at 10% frequency to decrease net surface hydrophobicity to mitigate aggregation and decrease the propensity for non-specific binders in the library.

Based on previous designs, we chose loop length variations of 5 to 7 residues in CDR H1 and 6 to 20 residues in CDR H3 while CDR H2 was kept constant at 17 residues (Kabat definitions) (**Extended Data Fig. 2.1, Supplementary Table 2.2**).[30,32] To cover this large sequence space with a minimal bias towards different length variants, five separate sub-libraries were constructed by binning CDR H3 loop length insertions ($X_{2\text{-}16}$) in incremental sets of three and combined to yield a final library of ~5 x 10^{10} transformants (**Supplementary Figure 2.1**). Analysis of the unique

CDR H3 sequences by next-generation sequencing (NGS) showed that observed amino acid frequencies closely matched our designs and all CDR H3 length variants were represented in the final library (**Fig. 2.1b-d, Supplementary Figure 2.2**). Finally, to test the performance of the library, several rounds of panning were performed on representative antigens including both cytosolic and membrane proteins. These panning experiments were done in parallel with an in-house Fab-phage library. For all target antigens, high levels of enrichment were observed (**Supplementary Figure 2.3**). For the majority of antigens, enrichment levels were comparable or substantially higher for the VH-phage library compared to the Fab-phage library.

Identification of VHs that target multiple epitopes

To date, most neutralizing mAbs against SARS-CoV-2 target Spike, and not surprisingly many of the most potent target the ACE2 binding interface.[3,7] While cryo-EM structures show that the ACE2 binding interface remains largely solvent-accessible in both the RBD "up" and "down" conformations,[33] simultaneous intra-molecular engagement by both binding arms of mAbs may be challenging as they are not geometrically arranged to engage multiple RBDs on a single Spike trimer. Thus, our goal was to target this highly neutralizing epitope with VHs and subsequently link them together to utilize avidity beyond that of a homo-bivalent mAb.

We first expressed the Spike-RBD (residues 328-533) and the ACE2 peptidase domain (residues 1-640) as biotinylated Fc-fusions for VH-phage selections.[34] To specifically enrich for VH-phage that bind the ACE2 binding site on Spike-RBD, the library was first cleared with the Spike-RBD-Fc/ACE2-Fc complex to remove phage that bound outside the ACE2 binding interface. This was followed by selection on Spike-RBD-Fc alone to enrich for phage that bound the unmasked ACE2 binding site **(Fig. 2.2a)**. By rounds 3 and 4, significant enrichments for phage

that bound Spike-RBD-Fc but not to Spike-RBD-Fc/ACE2-Fc complex were observed **Supplementary Figure 2.4a)**. Single clones were isolated and characterized for their ability to bind Spike-RBD-Fc by phage enzyme-linked immunosorbent assays (ELISA) **(Supplementary Figure 2.4b)**. Nearly all VH-phage showed enhanced binding to Spike-RBD-Fc over the Spike-RBD-Fc/ACE2-Fc complex, suggesting they bound the same epitope as ACE2 and could potentially block this interaction (**Supplementary Figure 2.4c**). In total, 85 unique VH-phage sequences were identified, and a subset was characterized as recombinant VH domains. We identified three lead VH candidates that bound Spike-RBD with K_D values ranging from 23-113 nM (**Fig. 2.2b-d**). These VH were specific to Spike-RBD and did not recognize other in-house antigens (**Extended Data Fig. 2.2**). Epitope binning demonstrated that the three VH domains mapped to two non-overlapping epitopes we call Site A and Site B within the larger ACE2 binding site (**Supplementary Figure 2.5**). The VH domain that recognizes Site A (A01) binds independently from the VHs that recognize Site B (B01 and B02) (**Fig. 2.2e-2h**).

Bi-paratopic and multivalent linkage increases affinity

We chose two parallel approaches to increase the affinity of the VH binders to Spike through avidity. First, we reasoned that VHs targeting Site A or Site B are in close proximity because they are non-overlapping but compete for the larger ACE2 binding site. Therefore, these VHs could be linked together to engage the same RBD simultaneously and improve affinity through intra-RBD avidity. Using the three VH monomers (A01, B01, B02) as modular units, we generated two bi-paratopic linked dimers (VH_2) by fusing A01 with B01 or B02 (**Fig. 2.3a**). In a parallel approach, we aimed to leverage the trimeric nature of Spike and engage multiple RBDs on the same Spike simultaneously to improve affinity through inter-RBD avidity. To that end, we

generated mono-paratopic Fc fusions (VH-Fc), linked dimers (VH_2), and linked trimers (VH_3) (**Fig. 2.3a**). The VH_2 and VH_3 consisted of a C-to-N terminal fusion of two or three VH monomers via a 20-amino acid Gly-Ser linker (~70 Å) while the VH-Fc consisted of a fusion of VH to the human IgG1 Fc domain via a flexible Fc hinge (~100 Å). The structure of the SARS-CoV-2 Spike trimer suggests the linked VH domains could bridge the distance between RBDs on an individual Spike (<55 Å), but are unlikely to span RBDs between discrete Spike proteins based on the inter-Spike distance on the viral envelope (150-180 Å) (**Extended Data Fig. 2.3**).[33,35]

ELISA and BLI binding assays to Spike-RBD show that the VH-Fc, VH_2, and VH_3 have 2.7 to 600-fold higher affinity to Spike-RBD (K_D = 0.1-8.4 nM) compared to the standalone VH monomers (**Fig. 2.3b-d, Extended Data Fig. 2.4, Supplementary Table 2.3 and 2.4, Supplementary Figure 2.6).** Interestingly, fold-increases in affinity were greater for binders that target Site B or both Site A and Site B combined. The most potent constructs bound trimeric Spike ectodomain (S_{ecto}) with K_Ds in the hundreds of picomolar range and all utilized VH B01 (**Fig. 2.3b-d**). Next, we examined whether these multivalent VH can block ACE2 binding to Spike by testing several high-affinity constructs (VH-Fcs; VH_2 A01-B01, VH_2 A01, and VH_2 B02) in a sequential BLI binding assay. S_{ecto} was immobilized on the biosensor, pre-blocked with each VH binder, and then assayed for binding to ACE2-Fc (**Fig. 2.3e**). We found that all binders tested substantially blocked binding of ACE2-Fc to S_{ecto}. Similarly, we examined whether these engineered VH-Fcs can compete with SARS-CoV-2 Spike-reactive antibodies in convalescent patient serum. Using a competition ELISA format previously developed by our group,[36] we found that VH-Fcs reduced the binding of patient antibodies to SARS-CoV-2 Spike-RBD (**Fig. 2.3f**). Taken together these data show that modular reformatting of these VH domains can significantly increase the affinity to Spike and block the same immunogenic epitopes as patient-derived Abs.

Lastly, we characterized the biophysical properties of these engineered VH by differential scanning fluorimetry (DSF), size exclusion chromatography (SEC), and reconstitution after lyophilization. The VH binders can be expressed in *E. coli* at high yields (i.e. VH_2 A01-B01 and VH_3 B01 express at ~1 g/L in shake flask culture) and have good stabilities (T_m = 56-65 °C) (**Supplementary Figure 2.7**). The most potent binders elute as a single monodisperse peak via SEC (**Supplementary Figure 2.8**), and VH_3 B01 retains binding to Spike-RBD and a monodisperse SEC profile after lyophilization and reconstitution suggesting it could be suitable for lyophilized formulation (**Supplementary Figure 2.9**).

Bi-paratopic and multivalent VH neutralize virus

SARS-CoV2 is a biosafety level 3 (BSL3) pathogen. To facilitate studies under routine laboratory conditions, we utilized a pseudotyped lentiviral (HIV) particle that has been previously reported.[37] These pseudotyped particles are generated from a three-plasmid system containing non-Env proteins from HIV, a luciferase reporter, and the SARS-CoV-2 spike protein. Entry of this viral particle into ACE2-expressing target cells and neutralization by anti-Spike antibodies have previously been shown to faithfully recapitulate features of the authentic pathogen without the need for working under BSL3 conditions. This pseudotyped virus was used to determine the half-maximal inhibitory concentration (IC_{50}) of neutralization for each construct.

The VH monomers neutralized pseudotyped virus weakly (IC_{50} > 50 nM), and cocktails of unlinked monomers did not improve potency. In contrast, the multivalent binders (VH_2, VH_3, and VH-Fc) neutralized ~10-1000 fold more potently compared to their respective monomeric units (**Fig. 2.4a, Supplementary Table 2.5, Extended Data Fig. 2.5**). There was a linear correlation

between the in vitro binding affinity (K_D) to Spike-RBD and the pseudotyped neutralization potency (IC_{50}) across the different binders ($R^2 = 0.72$) (**Fig. 2.4b**).

In particular, we observed that bi-paratopic VH_2 A01-B01 and VH_2 A01-B02 were stronger neutralizers than unlinked monomer cocktails. Additionally, the neutralization curves of the bi-paratopic (multi-site) VH_2 differed from the homodimeric (single-site) VH_2 binding to either Site A or B. That is, the bi-paratopic VH_2 exhibited a more cooperative transition and fully neutralized virus, while the homodimeric VH_2 showed a more linear transition and did not fully block viral entry even at high concentrations (**Fig. 2.4c**). This may reflect mechanistic differences; the bi-paratopic VH_2 can theoretically engage a single RBD using both VH domains simultaneously (intra-RBD avidity) and more fully occlude the ACE2 binding site, while the homodimeric VH_2 must bridge separate RBDs within the trimer (inter-RBD avidity).

Furthermore, the increase in neutralization potency as we increase the number of tandem VH units is exemplified by the VH B01-derived binders, as the IC_{50}s of VH_2 B01 and VH_3 B01 are two to three orders of magnitude lower than the IC_{50} of the VH B01 monomer (**Fig. 2.4d**). This is also observed for VH-Fc B01, which also neutralizes two orders of magnitude more potently than the monomer. Interestingly, although the neutralization potency of VH_2 A01 is better than VH A01, the potency does not improve further when a third domain is added (VH_3 A01), indicating that epitope-specific geometries can affect the extent to which increasing valency improves potency. The pseudotyped virus neutralization assays demonstrate that the top predicted binders from *in vitro* affinity data, VH_2 A01-B01, VH_2 A01-B02, VH_3 B01, and VH-Fc B01 are indeed the most potent with IC_{50}s of 0.74 nM, 1.08 nM, 0.156 nM, and 1.86 nM, respectively.

Lastly, we tested the ability of the most potent VH binders to neutralize authentic SARS-CoV-2 virus. As predicted, VH_3-B01 neutralized most potently. The VH-Fc B01, VH_2 A01-B01,

and VH_2 A01-B02 followed trends consistent with both the *in vitro* binding K_D an d pseudotyped virus IC_{50}. VH_3 B01, VH-Fc B01, VH_2 A01-B01, and VH_2 A01-B02 blocked authentic SARS-CoV-2 viral entry with IC_{50}s of 3.98 nM, 33.5 nM, 12.0 nM, and 26.2 nM, respectively (**Fig. 2.4e, Supplementary Table 2.6**).

Cryo-EM confirms multivalent binding of VH_3 B01 to Spike

To confirm whether our linking strategy could successfully engage multiple RBDs on Spike, we obtained a 3.2 Å global resolution cryo-EM 3D reconstruction of SARS-CoV-2 S_{ecto} in complex with VH_3 B01, the most potent neutralizer (**Fig. 2.5a, Extended Data Fig. 2.6, Supplementary Figure 2.10, Supplementary Table 2.7**). Although the S2 region of S_{ecto} was resolved at the reported resolution, the RBDs with the bound VH domains were resolved at about 6 Å resolution. However, even at this resolution the structure unambiguously revealed the three RBDs on Spike are in a two "up" and one "down" conformation. Densities corresponding to each VH domain are present on all three RBDs, indicating that VH B01 can bind both "up" and "down" conformations of RBD. SARS-CoV-2 Spike is rarely observed in this conformation with most structures being in all RBD "down" state or one RBD "up" state.[4,6,33,38] The binding epitope of VH B01 overlaps with the known ACE2 binding site (**Fig. 2.5b**), confirming the intended mechanism of neutralization and validating the ability of the masked selection strategy to precisely direct a binder toward the intended surface on a target protein. To further investigate the dependence of binding on Spike conformation, we performed BLI experiments with S_{ecto} at pH 4.5, which has been recently reported to lock Spike in an "all-down" RBD conformation.[39] Binding of both VH_3 B01 and VH_2 A01-B01 to S_{ecto} were greatly diminished (**Extended Data Fig. 2.7**). This

observation suggests that one or more RBDs in the "up" conformation may be required for VH B01 to access multiple RBDs on Spike.

2.4 Discussion

Here we describe a straightforward strategy to rapidly generate linked single-domain binders that potently neutralize SARS-CoV-2. We began by creating and validating a diverse human VH-phage library and generating VH binders to the ACE2 binding interface of Spike-RBD by a masked selection approach. From a panel of 85 unique VH binders, three were identified that recognized two separate epitopes within the ACE2 binding interface with nM affinity. In order to engage multiple epitopes simultaneously within an RBD or across RBDs on Spike, these VH monomers were linked into multivalent and bi-paratopic formats by Gly-Ser linkers or Fc domains without any further high-resolution structural information. This linkage approach not only significantly enhanced affinity, but also substantially improved the viral neutralization potency. We confirmed the basis of this increase by obtaining a cryo-EM structure of the most potent trivalent VH bound to Spike. Consistent with our original design, the trivalent VH simultaneously blocked the ACE2 binding site on all three RBDs of Spike.

We show that *in vitro* binding affinities and neutralization potencies against this oligomeric target can be dramatically increased with valency. This is exemplified by VH B01, which shows a 460-fold increase in binding affinity and 1400-fold increase in pseudotyped virus neutralization from VH B01 to VH_2 B01 to VH_3 B01. The cryo-EM structure of VH_3 B01 shows this binder neutralizes SARS-CoV-2 by blocking the ACE2 binding site on all three RBDs on Spike. Although there could be other contributing neutralization mechanisms such as viral aggregation induced by inter-Spike or inter-virion binding by the multivalent VH, the relatively short linker length (20

amino acids) between the VH domains would make these long-distance interactions significantly less likely. Additionally, no aggregates or S_{ecto} multimers induced by VH_3 B01 were observed during cryo-EM sample preparation and data collection, which further suggests that inter-Spike or inter-virion binding effects, if present, are minimal.

A similar relationship between valency and potency is observed for VH B02, which also targets the same epitope (Site B). However, VH A01, which targets Site A, is mechanistically distinct as there is no change in potency between a bivalent (VH_2) and a trivalent (VH_3) format. This suggests that unlike Site B binders, the trivalent Site A binder may not be able to fully engage all 3 RBDs. This could be due to the specific binding mechanism and epitope of VH A01, conformational differences of the RBDs within the Spike trimer, or spatial constraints of the linker. Structure determination of Site A binders in complex with Spike can elucidate the mechanistic and geometric difference between Site A and Site B epitopes. Additionally, structure-guided approaches to optimize linker lengths and orientations, coupled with an affinity maturation campaign may enable further increases in potency beyond what is demonstrated in this study.

Interestingly, we observed a difference in the cooperativity of the pseudotyped virus neutralization curves between bivalent VH_2 that target both Site A and Site B (bi-paratopic) versus VH_2 that target only Site A or Site B (mono-paratopic). Despite similar IC_{50}s, the IC_{95}s for bi-paratopic VH_2 (VH_2 A01-B01, VH_2 A01-B02) were much lower than mono-paratopic VH_2 binders. This could indicate a mechanistic difference between these two types of bi-valent binders. Bi-valent VH_2 can engage multiple RBD, but only 2 of the 3 RBDs can be efficiently occluded, leaving one RBD accessible to ACE2. This could underlie why the neutralization of mono-paratopic VH_2 does not reach 100% compared to say, a trivalent VH_3. Curiously, however, the bi-paratopic VH_2 can fully neutralize virus, despite also having only two VH domains. We reason

that intra-RBD avidity may play an important and unique role in the neutralization mechanism of bi-paratopic VH_2. Mono-paratopic binders are limited to inter-RBD avidity and regardless of valency are limited to up to 3 binding sites on Spike. In contrast, bi-paratopic VH_2 can utilize up to 6 binding epitopes on Spike and utilize both inter-RBD and intra-RBD avidity. Additionally, a more complete occlusion of the ACE2 binding interface (~864 $Å^2$)[40] on RBD by the bi-paratopic VH_2 could also underlie this difference in neutralization profiles. Although we do not have a structure of VH_2 A01-B01, we know both A01 and B01 can bind simultaneously within the ACE2 binding site. The cryo-EM structure for B01 shows good coverage of the ACE2 binding site (**Fig. 2.5B**) while leaving open adjacent space for VH A01 to occupy a non-overlapping epitope. A better understanding of the binding and neutralization mechanisms of bi-paratopic binders remain a current area of investigation and could lead to the engineering of more potent, multi-specific binders.

Previous studies that utilized designed ankyrin repeat proteins (DARPins), llama derived nanobodies, computationally designed proteins, and bivalent Fabs inspired our engineering strategy.[41–47] We believe our human-derived VH domains offer distinct advantages. Foremost, they would not require the time-consuming structure-guided humanization process that would be necessary for therapeutic nanobody development. Additionally, the VH domains derived from our library have favorable biophysical properties derived from a shared well-behaved scaffold. Our most potent construct, VH_3 B01, retains high-affinity binding after lyophilization cycles. This factor coupled with scalable production in bacterial systems may enable lung delivery via inhalation and could facilitate rapid deployment in response to a pandemic. Despite these advantages, single-domain binders have comparatively lower stability and *in vivo* half-life than mAbs, and due to their novelty, pharmacodynamic properties, such as the prevalence of anti-drug

antibodies, have not been as extensively tested in the clinic. Despite these challenges, the engineering potential of these single-domain binders is immense and provides an exciting avenue for the next generation of therapeutic antibody engineering and development. Our results illustrate this potential through a straightforward and rapid strategy to improve the efficacy of single-domain binders and could be applied to other protein interfaces of interest, including future viral proteins and other antigens associated with human disease.

Acknowledgements:

We thank members of the Wells Lab, particularly those working on COVID-19 projects for their efforts and contributions. Specifically, we would like to thank J. Gramespacher for purifying binders and M. Nix (UCSF) for guidance on pseudotyped virus neutralization assays. Additionally, we thank the entire QCRG for its rapid large-scale collaborative effort. Specifically, we thank C. Puchades and C. Azumaya for their efforts in the optimization of cryo-EM grid freezing and data collection; T. Owens for assistance in protein purification; D. Diwanji for expression of S_{ecto}; and A. Manglik (UCSF) for advice on cryo-EM experiments and protein purification. We also thank the laboratory of P. Kim (Stanford University) for providing plasmids for pseudotyped virus production. Lastly, we thank M. Wilson, C. Chiu, and R. Loudermilk (UCSF) as well as the patients, for providing convalescent sera.

J.A.W. is supported by generous grants from NCI (R35 GM122451-01); Chan-Zuckerberg Biohub, UCSF Program for Breakthrough Biomedical Research (PBBR); Fast Grants from Emergent Ventures at the Mercatus Center, George Mason University (#2154); and funding from The Harrington Discovery Institute (GA33116). S.A.L. is a Merck Fellow of the Helen Hay

Whitney Foundation. K.S. is a Fellow of the Helen Hay Whitney Foundation. X.X.Z is a Merck Fellow of the Damon Runyon Cancer Research Foundation, DRG-2297-17. J.R.B. and J.Z. are supported by a National Institutes of Health National Cancer Institute F32 5F32CA239417 (to J.R.B,) and 5F32CA236151-02 (to J.Z.). B.S.Z. is supported by a National Institutes of Health National Cancer Institute T32 (HL007185). N.J.R and I.L are supported by the National Science Foundation (GRFP). D.P.N was the Connie and Bob Lurie Fellow of the Damon Runyon Cancer Research Foundation (DRG-2204-14) and supported by a UCSF-PBBR Postdoctoral Independent Research Award, which is partially funded by the Sandler Foundation.

QCRG Structural Biology Consortium.

In addition to those listed explicitly in the author contributions, the structural biology portion of this work was performed by the QCRG (Quantitative Biosciences Institute Coronavirus Research Group) Structural Biology Consortium. Listed below are the contributing members of the consortium listed by teams in order of team relevance to the published work. Within each team the team leads are italicized (responsible for organization of each team, and for the experimental design utilized within each team), then the rest of team members are listed alphabetically. **CryoEM grid freezing/collection team:** *Caleigh M. Azumaya, Cristina Puchades, Ming Sun*, Julian R. Braxton, Axel F. Brilot, Meghna Gupta, Fei Li, Kyle E. Lopez, Arthur Melo, Gregory E. Merz, Frank Moss, Joana Paulino, Thomas H. Pospiech, Jr., Sergei Pourmal, Alexandrea N. Rizo, Amber M. Smith, Paul V. Thomas, Feng Wang, Zanlin Yu. **CryoEM data processing team:** *Miles Sasha Dickinson, Henry C. Nguyen*, Daniel Asarnow, Julian R. Braxton, Melody G. Campbell, Cynthia M. Chio, Un Seng Chio, Devan Diwanji, Bryan Faust, Meghna Gupta, Nick Hoppe, Mingliang Jin, Fei Li, Junrui Li, Yanxin Liu, Gregory E.

Merz, Joana Paulino, Thomas H. Pospiech, Jr., Sergei Pourmal, Smriti Sangwan, Raphael Trenker, Donovan Trinidad, Eric Tse, Kaihua Zhang, Fengbo Zhou. **Mammalian cell expression team:** *Christian Billesboelle, Melody G. Campbell, Devan Diwanji, Carlos Nowotny, Amber M. Smith, Jianhua Zhao*, Caleigh M. Azumaya, Alisa Bowen, Nick Hoppe, Yen-Li Li, Phuong Nguyen, Cristina Puchades, Mali Safari, Smriti Sangwan, Kaitlin Schaefer, Raphael Trenker, Tsz Kin Martin Tsui, Natalie Whitis. **Protein purification team**: *Daniel Asarnow, Michelle Moritz, Tristan W. Owens, Sergei Pourmal*, Caleigh M. Azumaya, Cynthia M. Chio, Bryan Faust, Meghna Gupta, Kate Kim, Joana Paulino, Jessica K. Peters, Kaitlin Schaefer, Tsz Kin Martin Tsui. **Crystallography team:** *Nadia Herrera, Huong T. Kratochvil, Ursula Schulze-Gahmen, Iris D. Young*, Justin Biel, Ishan Deshpande, Xi Liu. **Bacterial expression team**: *Amy Diallo, Meghna Gupta*, *Erron W. Titus*, Jen Chen, Loan Doan, Sebastian Flores, Mingliang Jin, Huong T. Kratochvil, Victor L. Lam, Yang Li, Megan Lo, Gregory E. Merz, Joana Paulino, Aye C. Thwin, Zanlin Yu, Fengbo Zhou, Yang Zhang. **Infrastructure team:** David Bulkley, Arceli Joves, Almarie Joves, Liam McKay, Mariano Tabios, Eric Tse. **Leadership team:** *Oren S Rosenberg, Kliment A Verba*, David A Agard, Yifan Cheng, James S Fraser, Adam Frost, Natalia Jura, Tanja Kortemme, Nevan J Krogan, Aashish Manglik, Daniel R. Southworth, Robert M Stroud. The QCRG Structural Biology Consortium has received support from: Quantitative Biosciences Institute, Defense Advanced Research Projects Agency HR0011-19-2-0020 (to D.A.A. and K.A.V.; B. Shoichet PI), FastGrants COVID19 grant (K.A.Verba PI), Laboratory For Genomics Research (O.S. Rosenberg PI) and Laboratory For Genomics Research (R.M. Stroud PI)

2.5 Materials and Methods

VH Library Construction and Validation by NGS

The VH-phage library was created through bivalent display of VH on the surface of M13 bacteriophage as has been previously described.[30,48,49] In brief, the DNA phagemid library was created through oligonucleotide mutagenesis. First, the human VH-4D5 sequence was modified with five mutations (35G/39R/45E/47L/50R) in the framework and with restriction sites in each of the CDRs: AgeI in CDR H1, NcoI in CDR H2, and XhoI in CDR H3 (**Supplementary Table 2.1**). Oligonucleotides were synthesized by a custom Trimer Phosphoramidite mix for CDR H1 and CDR H2 (Twist Bioscience) and CDR H3 (Trilink Biotechnologies, Inc.) (**Supplementary Table 2.2**). After mutagenesis DNA sublibrary pools were digested with appropriate restriction enzymes to remove the phagemid template before transformation into SS320 electrocompetent cell (Lucigen) for phage production. NGS of the CDR H3 was performed on the pooled library by amplifying the phagemid from boiled phage with in-house primers. Samples were submitted for analysis on a HiSeq4000 (Illumina) with a custom primer: TGAGGACACTGCCGTCTATTATTGTGCTCGC (T_m = 67 °C, GC% = 52). NGS analysis of output was performed using an in-house informatics pipeline written in R. In brief, the raw NGS data sequencing file (*.fastq.gz) was converted into a table comprised of the DNA sequences, the amino acid sequences (CDR H3), and the counts/frequency as columns and then saved as a *.csv file for further analysis (e.g., calculation of: amino acid abundancy, sequence logo, H3 length distribution, etc.). Several filters were applied: i) low-quality sequences containing "N" were removed, ii) sequences with any stop codon were removed; iii) only the sequences that were in-frame were kept. Scripts are available for download at: https://github.com/crystaljie/VH_library_CDR_H3_NGS_analysis_Cole.Bracken.git.

Cloning, Protein Expression, and Purification

Spike-RBD-monomer, Spike-RBD-Fc, Spike ectodomain (S_{ecto}), and ACE2-Fc were produced as biotinylated proteins as previously described.[34] VH were subcloned from the VH-phagemid into an *E. coli* expression vector pBL347. VH_2 and VH_3 were cloned into pBL347 with a 20-amino acid Gly-Ser linker. VH-Fc were cloned into a pFUSE (InvivoGen) vector with a human IgG1 Fc domain. All constructs were sequence verified by Sanger sequencing. VH, VH_2, and VH_3 constructs were expressed in *E. coli* C43(DE3) Pro + using an optimized autoinduction media and purified by protein A affinity chromatography similarly to Fabs (**Supplementary Figure 2.11**).[30] VH-Fc were expressed in Expi293 BirA cells using transient transfection (Expifectamine, Thermo Fisher Scientific). Four days after transfection, the media was harvested, and VH-Fc were purified using protein A affinity chromatography. All proteins were buffer exchanged into PBS by spin concentration and stored in aliquots at -80ºC. The purity and integrity of proteins were assessed by SDS-PAGE. All proteins were endotoxin removed using an endotoxin removal kit (Thermo Fischer) prior to use in neutralization assays.

Phage selection with VH-phage library

Phage selections were done according to previously established protocols.[30] Selections were performed using biotinylated antigens captured with streptavidin-coated magnetic beads (Promega). In each round, the phage pool was first cleared by incubation with beads loaded with 500 nM ACE2-Fc/Spike-RBD-Fc complex. The unbound phage were then incubated with beads loaded with Spike-RBD-Fc. After washing, the bound phage was eluted by the addition of 2 μg/mL of TEV protease. In total, four rounds of selection were performed with decreasing amounts of

Spike-RBD-Fc as indicated in Figure 2A. All steps were done in PBS buffer + 0.02% Tween-20 + 0.2% BSA (PBSTB). Individual phage clones from the third and fourth round of selections were analyzed by phage ELISA.

Phage ELISA

For each phage clone, 4 different conditions were tested—Direct: Spike-RBD-Fc, Competition: Spike-RBD-Fc with an equal concentration of Spike-RBD-Fc in solution, Negative selection: ACE2-Fc/Spike-RBD-Fc complex, and Control: Fc. 384-well Nunc Maxisorp flat-bottom clear plates (Thermo Fisher Scientific) were coated with 0.5 μg/mL of NeutrAvidin in PBS overnight at 4ºC and subsequently blocked with PBS + 0.02% Tween-20 + 2% BSA for 1 hr at room temperature. Plates were washed 3X with PBS containing 0.05% Tween-20 (PBST) and were washed similarly between each of the steps. 20 nM of biotinylated Spike-RBD-Fc, ACE2-Fc/Spike-RBD-Fc complex, or Fc diluted in PBSTB was captured on the NeutrAvidin-coated wells for 30 min, then blocked with PBSTB + 10 μM biotin for 30 min. Phage supernatant diluted 1:5 in PBSTB were added for 20 min. For the competition samples, the phage supernatant was diluted into PBSTB with 20 nM Spike-RBD-Fc. Bound phage were detected by incubation with anti-M13-HRP conjugate (Sino Biological)(1:5000) for 30 min, followed by the addition of TMB substrate (VWR International). The reaction was quenched with the addition of 1 M phosphoric acid and the absorbance at 450 nm was measured using a Tecan M200 Pro spectrophotometer.

ELISA EC50 with Spike-RBD-monomer

384-well Nunc Maxisorp flat-bottom clear plates were prepared similarly to the phage ELISA protocol (above) by coating with neutravidin, followed by blocking with PBST + 2% BSA,

incubation with 20 nM Spike-RBD-monomer, and blocking by PBSTB + 10 μM biotin. VH binders in 4-fold dilutions ranging from 500 nM to 2.8 pM were added for 1 hour. Bound VH was detected by incubation with Protein A HRP conjugate (Thermo Fischer Scientific) (1:10,000) for 30 min, followed by the addition of TMB substrate for 5 min, quenching by 1 M phosphoric acid, and detection of absorbance at 450 nm. Each concentration was tested in duplicate, and the assay was repeated three times.

Bio-layer interferometry (BLI) Experiments

Bio-layer interferometry data (BLI) were measured using an Octet RED384 (ForteBio) instrument. Spike-RBD or S_{ecto} were immobilized on a streptavidin or Ni-NTA biosensor and loaded until a 0.4 nm signal was achieved. After blocking with 10 μM biotin, purified binders in solution was used as the analyte. PBSTB was used for all buffers for BLI at pH 7.4. For BLI at pH 4.5, 10 mM sodium acetate pH 4.5, 150 mM NaCl, 0.2 mg/mL BSA and 0.01% (w/v) Tween-20 was used. Data were analyzed using the ForteBio Octet analysis software and kinetic parameters were determined using a 1:1 monovalent binding model.

Competition ELISA with COVID-19 convalescent patient sera

Competition ELISA with convalescent patient sera was conducted with the same patient sera as previously reported.[36] Samples were collected in accordance with the Declaration of Helsinki using protocols approved by the UCSF Institutional Review Board (Protocol 20-30338). Patients were voluntarily recruited based on their history of prior SARS-CoV-2 infection. All patients provided written consent. Patient sera were de-identified prior to delivery to the Wells Lab, where all experiments presented here were performed. Briefly, sera were obtained as

described from patients with a history of positive nasopharyngeal SARS-CoV-2 RT-PCR test and at least 14 days after the resolution of their COVID-19 symptoms. Healthy control serum was obtained prior to the emergence of SARS-CoV-2. Sera were heat-inactivated (56°C for 60 min) prior to use. Competitive serology using biotinylated SARS-CoV-2 Spike-RBD as the capture antigen was performed as previously reported with slight modifications.[36] Instead of supplementing sera diluted 1:50 in 1% nonfat milk with 100 nM ACE2-Fc, 100 nM of each of the indicated VH-Fc fusions was used. Bound patient antibodies were then detected using Protein L-HRP (Thermo Fisher Scientific 32420, [1:5000]). Background from the raw ELISA signal in serum-treated wells was removed by first subtracting the signal measured in NeutrAvidin-alone coated wells then subtracting the signal detected in antigen-coated wells incubated with 1% nonfat milk + 100 nM competitor.

Differential Scanning Fluorimetry (DSF)

DSF was conducted as previously described.[30] Briefly, purified protein was diluted to 0.5 μM or 0.25 μM in buffer containing Sypro Orange 4x (Invitrogen) and PBS and assayed in a 384-well white PCR plate. All samples were tested in duplicate. In a Roche LC480 LightCycler, the sample was heated from 30°C to 95°C with a ramp rate of 0.3°C per 30 sec and fluorescence signal at 490 nm and 575 nm were continuously collected. T_m was calculated using the Roche LC480 LightCycler software.

Size Exclusion Chromatography (SEC)

SEC analysis was performed using an Äkta Pure system (GE Healthcare) using a Superdex 200 Increase 10/300 GL column. 100 μL of 2-3 mg/mL of each analyte was injected and run with

a constant mobile phase of degassed 10 mM Tris pH 8.0 200 mM NaCl. Absorbance at 280 nm was measured. The post-lyophilization and reconstitution SEC was performed using an Agilent HPLC 1260 Infinity II LC System using an AdvanceBio SEC column (300 Å, 2.7 μm, Agilent). Fluorescence (excitation 285 nm, emission 340 nm) was measured.

Preparation of SARS-CoV-2 pseudotyped virus and HEK-ACE2 overexpression cell line

HEK293T-ACE2 cells were a gift from Arun Wiita's laboratory at the University of California, San Francisco. Cells are cultured in D10 media (DMEM + 1% Pen/Strep + 10% heat-inactivated FBS). Plasmids to generate pseudotyped virus were a gift from Peter Kim's lab at Stanford University and SARS-Cov-2 pseudotyped virus was prepared as previously described.[37] Briefly, plasmids at the designated concentrations were added to OptiMEM media with FuGENE HD Transfection Reagent (Promega) at a 3:1 FuGENE:DNA ratio, incubated for 30 min, and subsequently transfected into HEK-293T cells. After 24 hrs, the supernatant was removed and replaced with D10 culture media. Virus was propagated for an additional 48 hrs, and the supernatant was harvested and filtered. Virus was stored at 4ºC for up to 10 days.

HEK-ACE2 were seeded at 10,000 cells/well on 96-well white plates (Corning, cat. 354620). After 24 hrs, pseudotyped virus stocks were titered via a two-fold dilution series in D10 media, and 40 μL were added to cells. After 60 hrs, infection and intracellular luciferase signal was determined using Bright-Glo™ Luciferase Assay (Promega), and the dilution achieving maximal luminescent signal within the linear range, ~3-5 x 10^5 luminescence units, was chosen as the working concentration for neutralization assays. Pseudovirus stocks were flash-frozen in aliquots and stored at -80ºC and thawed on ice just prior to use in a neutralization assay.

Pseudotyped viral neutralization assays

HEK-ACE2 were seeded at 10,000 cells/well in 40 µL of D10 on 96-well white plates (Corning, cat. 354620) 24 hours prior to infection. To determine IC_{50} for pseudotyped virus, dose series of each VH binder were prepared at 3x concentration in D10 media and 50 µL were aliquoted into each well in 96-well plate format. Next, 50 µL of virus diluted in D10 media were added to each well and the virus and blocker solution was allowed to incubate for 1 hr at 37ºC. Subsequently, 80 µL of the virus and blocker solution were transferred to wells seeded with HEK-ACE2. After 60 hrs of infection at 37ºC, intracellular luciferase signal was measured using the Bright-Glo™ Luciferase Assay. 80 µL of reconstituted Bright-Glo™ luciferase reagent was added to each well, incubated at room temperature with gentle shaking for 5 min, before the luminescence was measured on a Tecan M200 pro spectrophotometer. The half-maximal inhibitory concentration (IC_{50}) was determined using a 4-parameter nonlinear regression (GraphPad Prism).

Authentic Viral Neutralization Assays

Authentic virus neutralization assays were done as previously described.[34] All handling and experiments using live SARS-CoV-2 virus clinical isolate 2019-nCoV/USA-WA1/2020 (BEI Resources) was conducted under Biosafety Level 3 containment with approved BUA and protocols. VeroE6 cells were cultured in Minimal Essential Media (MEM), 10% FBS, 1% Pen-Strep. For neutralization assays, VeroE6 cells were seeded on 6-well culture plates at 3.8E5 cells/well the day prior. Infection with SARS-CoV-2 was performed using MOI of 0.1. Virus was incubated in infection media (EMEM 0% FBS) containing different concentrations of binders for 1 hr at 37°C. Culture media was removed from VeroE6 cells and 300 µL of the blocker/virus

inoculum was added to cells for 1 hr at 37°C. After this step, 1 mL of EMEM with 10% FBS was added to the cells, and the cells were incubated 37°C for an additional 16 hrs before RNA harvest. Viral entry into cells and cellular transcription of viral genes was measured by qPCR using the N gene and host GUSB and host ACTB as controls. RNAeasy RNA extraction kit (Qiagen) was used for RNA extraction and Quantitect Reverse-transcriptase kit (Qiagen) was used to generate cDNA. qPCR reactions were prepared using SYBR Select Master Mix (Thermo). N gene and hGUSB gene primer concentration was 400 nM and annealing temperature was 58°C. Primer sequences (IDT) were the following – viral N gene: N_F = CACATTGGCACCCGCAATC; N_R = GAGGAACGAGAAGAGGCTTG; host gene: hGUSB_F = CTCATCTGGAATTTTGCCGATT; hGUSB_R =CCGAGTGAAGACCCCCTTTTTA. Relative copy number (RCN) of viral transcript level compared to host transcript was determined using the ΔΔCT method. The half-maximal inhibitory concentration (IC_{50}) was determined using a 4-parameter nonlinear regression (GraphPad Prism).

Expression and purification of Spike ectodomain for cryo-EM

To obtain pre-fusion spike ectodomain, methods similar to the previous reports were used.[33,50] The expression plasmid, provided by the McLellan lab, was used in a transient transfection with 100 mL, high-density Chinese Hamster Ovary (ExpiCHO, Thermo Fisher) culture following the "High Titer" protocol provided by Thermo Fisher. Six to nine days post-transfection, the supernatant was collected with centrifugation at 4,000xg at room temperature. The clarified supernatant was then incubated with Ni-Sepharose Excel resin (Cytiva Life Sciences) for ninety minutes at room temperature. After incubation, the nickel resin was washed with 20 mM Tris (pH 8), 200 mM NaCl, and 20 mM imidazole with ten column volumes. Protein was eluted

from the nickel resin with 20 mM Tris (pH 8), 200 mM NaCl, and 500 mM imidazole. Eluate was then concentrated with a 50 MWCO Amicon Ultra-15 centrifugal unit by centrifugation at 2500xg, room temperature. The eluate was concentrated, filtered with a 0.2 μm filter, and injected onto a Superose6 10/300 GL column equilibrated with 10 mM Tris (pH 8), 200 mM NaCl. The fractions corresponding to monodisperse spike were collected and the concentration was determined using a nanodrop.

Cryo-EM sample preparation, data collection and processing

2 μM Spike ectodomain was mixed with 5-fold excess VH_3 B01 and applied (3 μL) to holey carbon Au 200 mesh 1.2/1.3 Quantifoil grids. Grids were blotted and plunge frozen using Mark 4 Vitrobot (ThermoFisher) at 4°C and 100% humidity, utilizing blot force 0 and blot time of 4 sec. 1656 images were collected on Titan Krios (ThermoFisher) equipped with K3 direct detector operated in CDS mode (Gatan Ametek) and an energy filter (Gatan Ametek) at nominal magnification of 105000x (0.834 Å/physical pixel). Dose fractionated movies were collected with a total exposure of 6 seconds and 0.04 seconds per frame at a dose rate of 9 electrons per physical pixel per second. Movies were corrected for motion and filtered to account for electron damage utilizing MotionCor2.[51] Drift corrected sums were imported into cryoSPARC2 processing package.[52] Micrographs were manually curated, CTF was estimated utilizing patches and particles were picked with a Gaussian blob. Previous Spike ectodomain structure was imported as an initial model (low pass filtered to 30 Å) and multiple rounds of 3D and 2D classification were performed. Images were re-picked with the best looking 2D class averages low pass filtered to 30 Å, and multiple rounds of 3D classification were performed again to obtain a homogeneous stack of Spike trimer particles. Majority of the particles went into classes putatively representing excess unbound

VH_3 B01 and the final Spike like particle stack only contained ~21000 particles. Non-uniform homogeneous refinement of the particle stack resulted in global resolution of 3.2 Å (masked) utilizing 0.143 FSC cut off.[53] PDB:6X2B was rigid body fit into the resulting reconstruction in UCSF Chimera.[54] The RBDs of two Spike trimers were moved as a rigid body to accommodate the cryo-EM density. The cryo-EM reconstruction was low pass filtered to 6 Å to better visualize the VH densities. Homology model was built based on the PDB:4G80 for the VH domains and the resulting and individual VH domains were rigid body fit into the 6 Å cryo-EM density as depicted in **Fig. 5A**. The resulting model was relaxed into the cryo-EM map low pass filtered to 6 Å with Rosetta FastRelax protocol. For **Fig. S13** the local resolution was estimated using ResMap.[55] The final figures were prepared using ChimeraX.[56]

Data availability

Cryo-EM structural data have been deposited in the Protein Data Bank and Electron Microscopy Data Bank, (PDB 7JWB and EMD-22514). The cryo-EM structure of Spike trimer from PDB:6VSB and ACE2 from PDB: 6M17 and PDB:6M0J were used to for distance estimates. The cryo-EM structure of Spike trimer from PDB: 6X2B and crystal structure of a VH domain from PDB: 3P9W were used for cryo-EM data processing. Source data for Figures 1-4 and Extended Data Figures (2,4,5, and 7) are provided with this paper. Any additional information is available upon request.

Code availability

Code used to analyze NGS data has been deposited onto Github and is available for download at: https://github.com/crystaljie/VH_library_CDR_H3_NGS_analysis_Cole.Bracken.git.

2.6 Main Figures

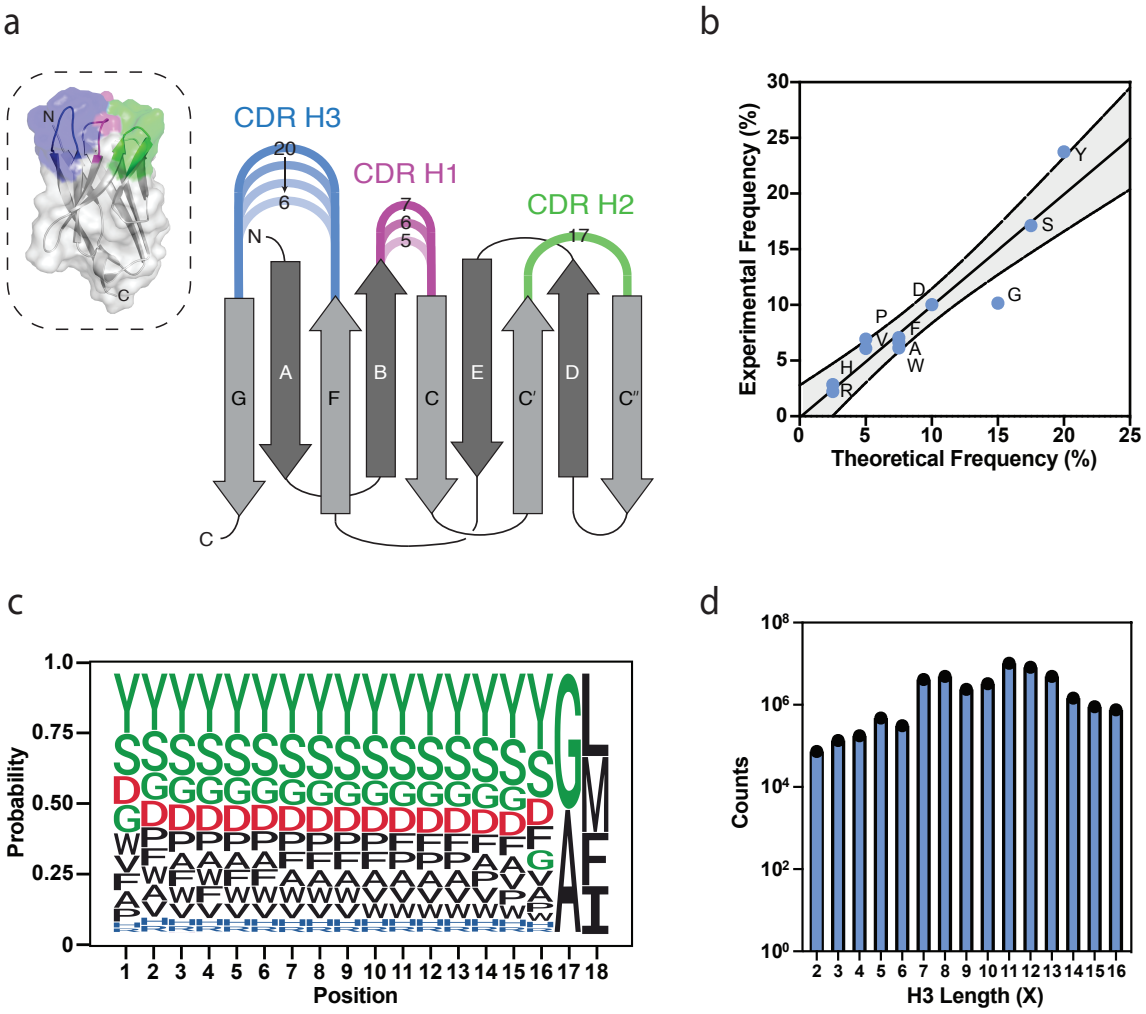


Figure 2.1: Design and validation of VH-phage library
(**A**) 3D surface representation (left) of the VH-4D5 parental scaffold (PDB:1FVC) and a cartoon diagram (right) where individual CDRs are annotated in color with the designed loop length variations according to Kabat nomenclature.[32] (**B**) NGS analysis of the longest H3 loop (X=16) shows that expected global amino acid frequencies are comparable to designed frequencies. Gray region denotes the 95% confidence interval. (**C**) Representative NGS analysis of the longest H3 loop (X=16) shows positional frequency distribution matches designed frequencies. Position 1 refers to residue 95 (Kabat definition). Data for the other CDR H3 lengths are reported in **Fig. S2**. (**D**) NGS analysis of unique clones shows that all H3 lengths are represented in the pooled VH-phage library.

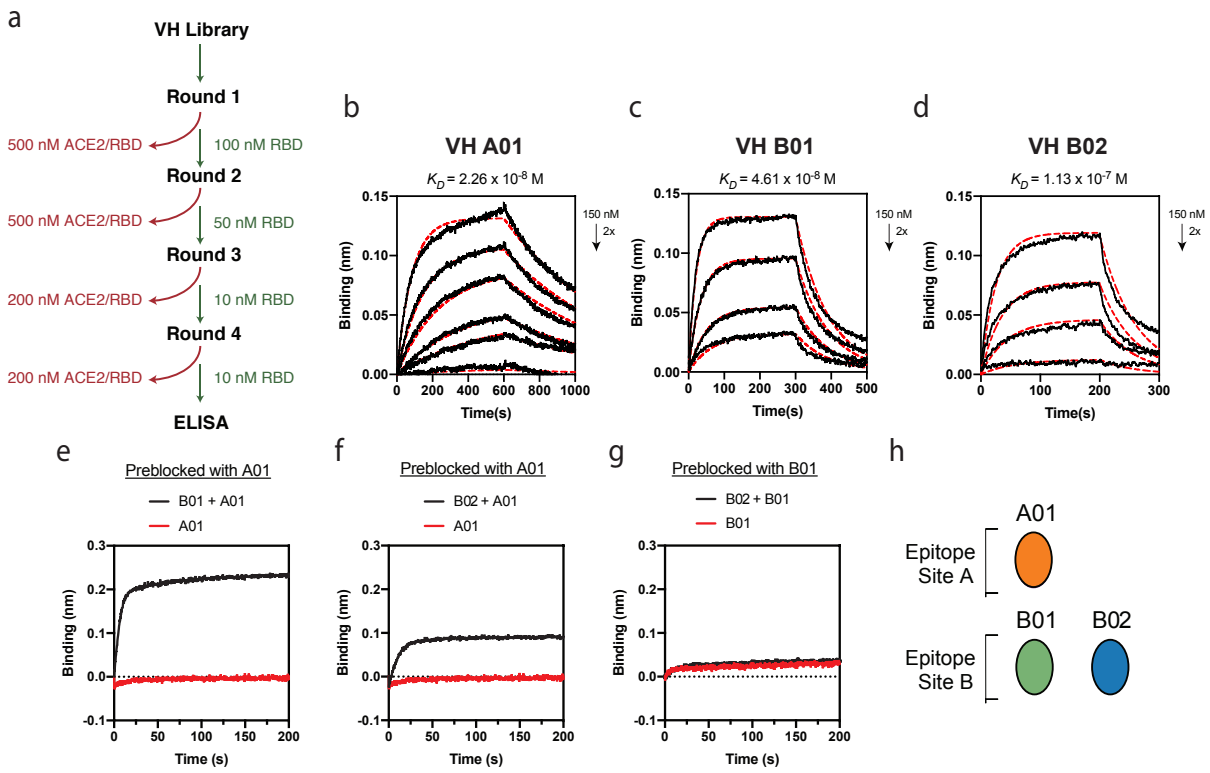


Figure 2.2: Identification of VH domains that bind Spike-RBD at two unique epitopes by phage display

(**A**) Diagram illustrating phage selection strategy to isolate VH-phage that bind at the ACE2 binding interface. Red indicates clearance of the phage pool by Spike-RBD-Fc/ACE2-Fc complex, green indicates positive selection against Spike-RBD-Fc alone. To increase stringency, successively lower concentrations of Spike-RBD-Fc were used, and after 4 rounds of selection, individual phage clones were analyzed by phage ELISA. BLI of (**B**) VH A01 (**C**) VH B01 and (**D**) VH B02 against Spike-RBD. (**E**) BLI-based epitope binning of VH A01 and VH B01, (**F**) VH A01 and VH B02, (**G**) VH B01 and VH B02. The antigen loaded onto the sensor tip was Spike-RBD. (**H**) Diagram of the two different epitope bins targeted by VH domains.

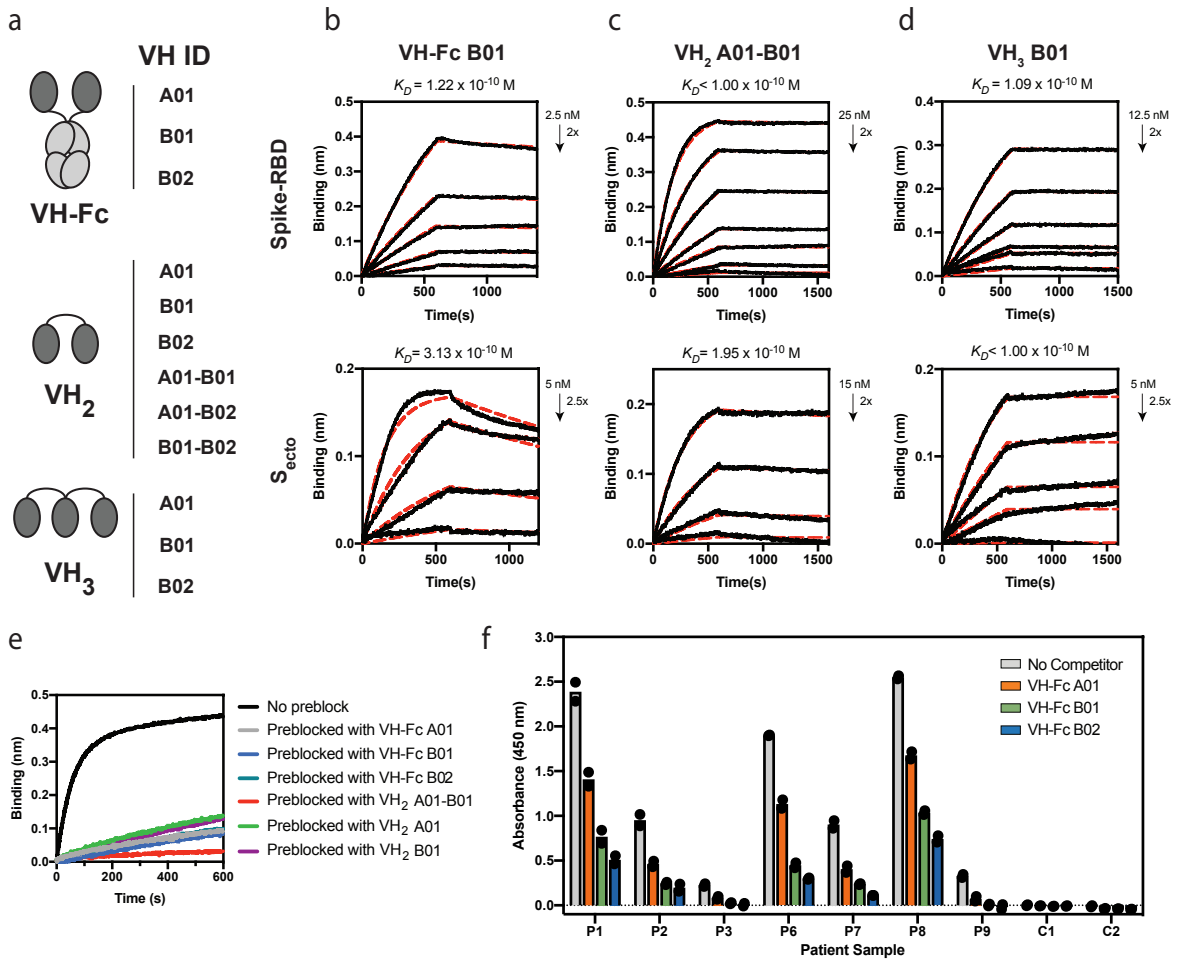


Figure 2.3: In vitro characterization of multivalent and bi-paratopic VH binders
(**A**) Cartoon depiction of engineered VH binders generated by linking VH domains via Fc-fusion or a 20-aa Gly-Ser linker. BLI traces of lead VH binders, (**B**) VH-Fc B01, (**C**) VH_2 A01-B01, (**D**) VH_3 B01 against RBD (upper panel) or S_{ecto} (lower panel). (**E**) Sequential BLI binding experiments that measured binding of ACE2-Fc to S_{ecto} pre-blocked with our VH binders show that multivalent VH binders can block ACE2-Fc binding to S_{ecto}. (**F**) Competition serology ELISA with convalescent patient sera indicates that VH-Fc binders can compete with patient antibodies. P1-P9 are sera from patients with a history of prior SARS-CoV-2 infection. C1-C2 are two donor sera collected before the SARS-CoV-2 outbreak. Individual data points represent technical replicates (n=2) from the serum of the same patient and are shown as black circles.

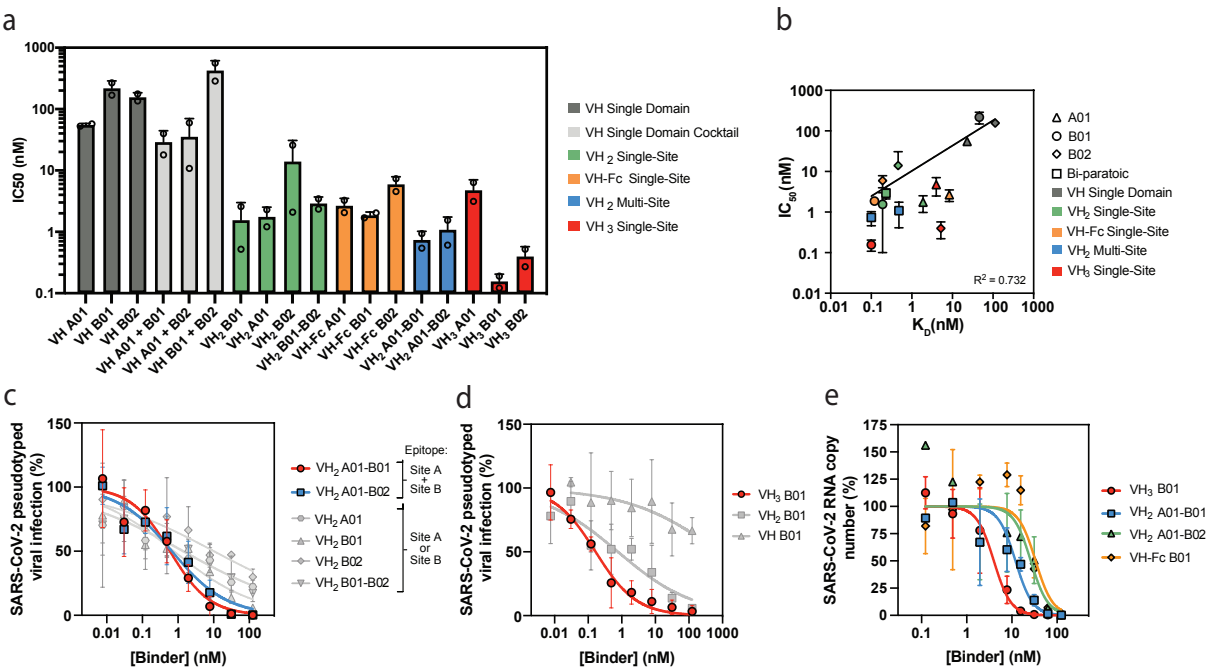


Figure 2.4: Multivalent and bi-paratopic VH binders neutralize pseudotyped and authentic SARS-CoV-2

(**A**) Pseudotyped virus IC_{50} of VH binders. Neutralization potency improves when VH domains are engineered into multivalent and bi-paratopic constructs. (**B**) Correlation of in vitro binding affinity (K_D) and pseudotyped virus neutralization (IC_{50}) of VH binders. Data were fit to a log-log linear extrapolation. (**C**) Pseudotyped virus neutralization curves of multi-site VH_2 in comparison to single-site VH_2 demonstrate that the multi-site VH_2 have a more cooperative neutralization curve. (**D**) Pseudotyped virus neutralization curves of mono-, bi-, and tri-valent formats of VH B01 demonstrate potency gains driven by valency. (**E**) Authentic SARS-CoV-2 neutralization curves for the most potent VH formats were determined via qPCR of viral genome in cellular RNA. All pseudoviral neutralization data were repeated as n=2 independent replicates. Authentic virus neutralization data were repeated as n=2 independent replicates. Data represent the average and standard deviation of replicates.

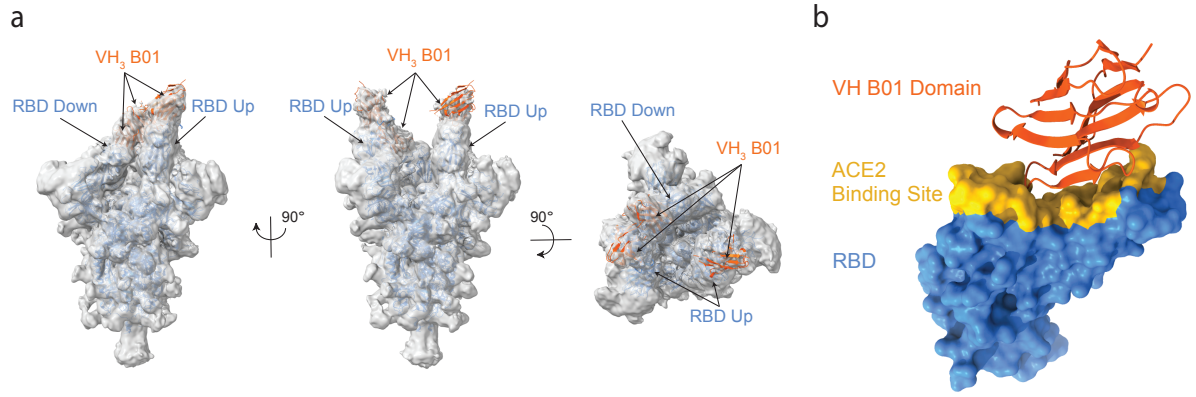


Figure 2.5: Cryo-EM reveals trivalent VH binding at the ACE2 binding interface of RBD (**A**) Side and top views of cryo-EM 3D reconstructions of VH_3 B01 + S_{ecto} (PDB: 7JWB) are shown with individual VH domain densities of VH_3 B01 fit with PDB: 3P9W (VH scaffold; orange cartoon). A total of three VH domains, each bound to an RBD of the Spike trimer, are resolved. 3D model of S_{ecto} was fit with reference structure (PDB:6X2B with additional rigid body fit of the individual RBDs; blue cartoon) and shows RBDs in a distinct two "up", one "down" conformation. Cryo-EM map was low-pass filtered to 6 Å. (**B**) View of the epitope (Site B) of one VH domain from VH_3 B01. Site B overlays directly with the ACE2 binding site (yellow surface; contacts defined as RBD residues within 8 Å of an ACE2 residue from PDB:6M0J).

2.7 Extended and Supplemental Figures

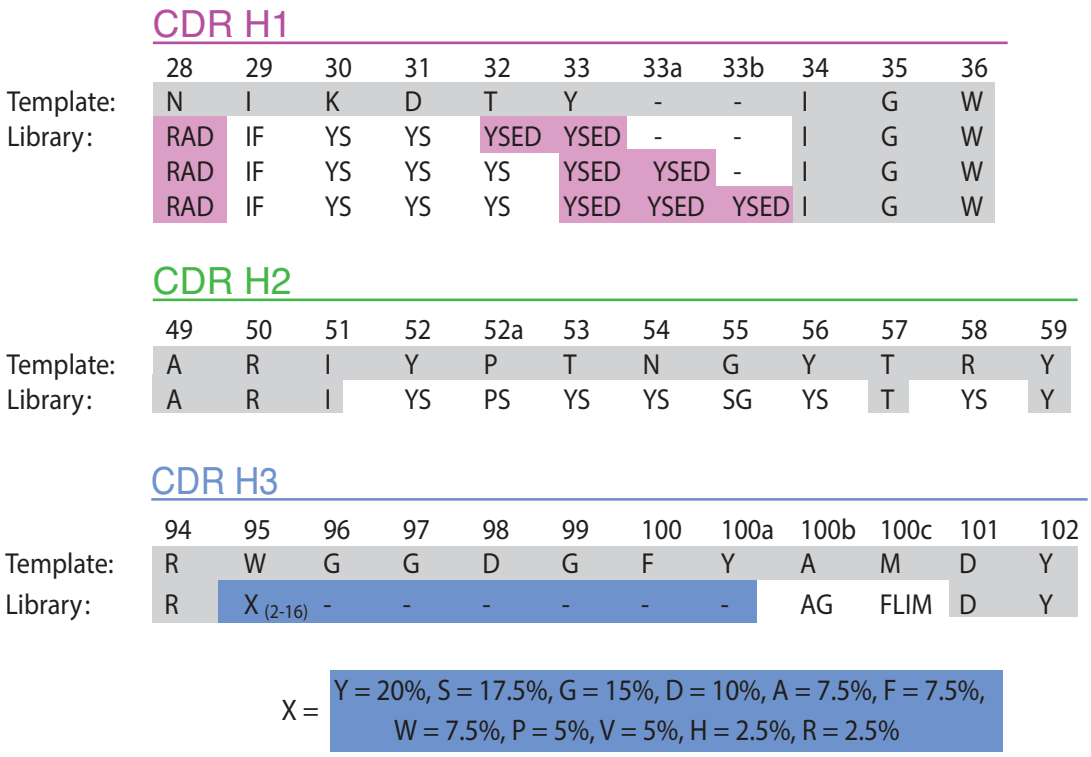


Figure 2.ED1
Schematic of CDR amino acid composition as compared to parental template. Positions in pink highlight CDR H1 charged amino acid insertions. Positions in blue highlight the insertion of 'X' synthetic amino acid mixture. Positions in gray remain unchanged from template.

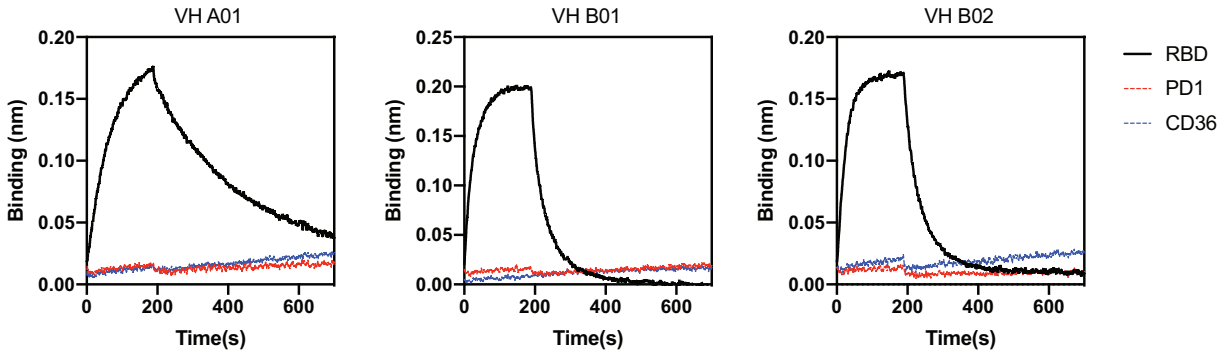


Figure 2.ED2
Bio-layer interferometry (BLI) binding traces of VH domains at 150 nM to RBD and two decoy antigens expressed as biotinylated Fc-fusions: programmed cell death protein 1 (PD-1), and platelet glycoprotein 4 (CD36). Respective VH show successful binding to RBD, but not to either of the decoy antigens.

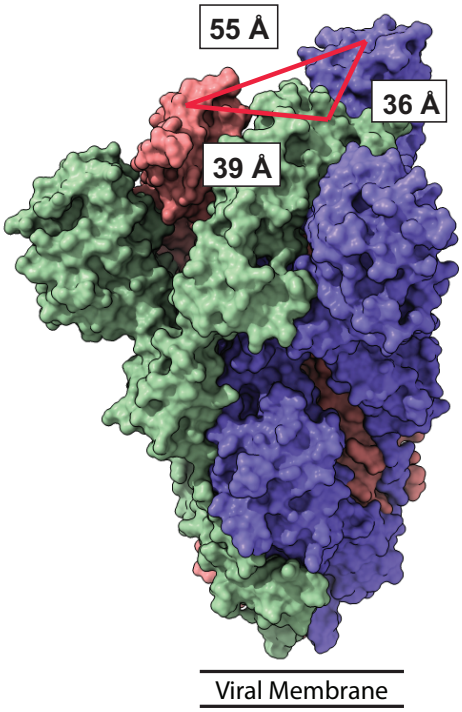


Figure 2.ED3
Structure of SARS-CoV-2 Spike trimer (PDB: 6VSB)32. RBD colored in blue is in the 'up' position while RBDs colored in green and pink are in the 'down' position. Distance between the mid-points of the ACE2 binding interface (PDB: 6M17)1 on respective RBDs was measured in Pymol.

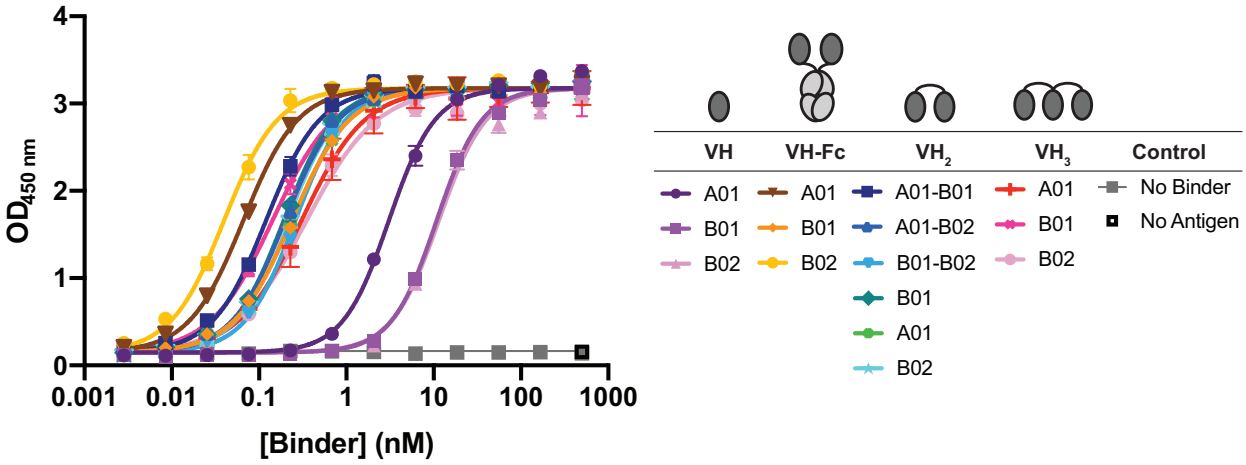


Figure 2.ED4
ELISA of VH binders against Spike-RBD. Data presented show the average and standard deviation from three independent experiments. Data were fit to a non-linear, four-parameter variable slope regression model using Prism 8 to obtain EC50 values for each binder.

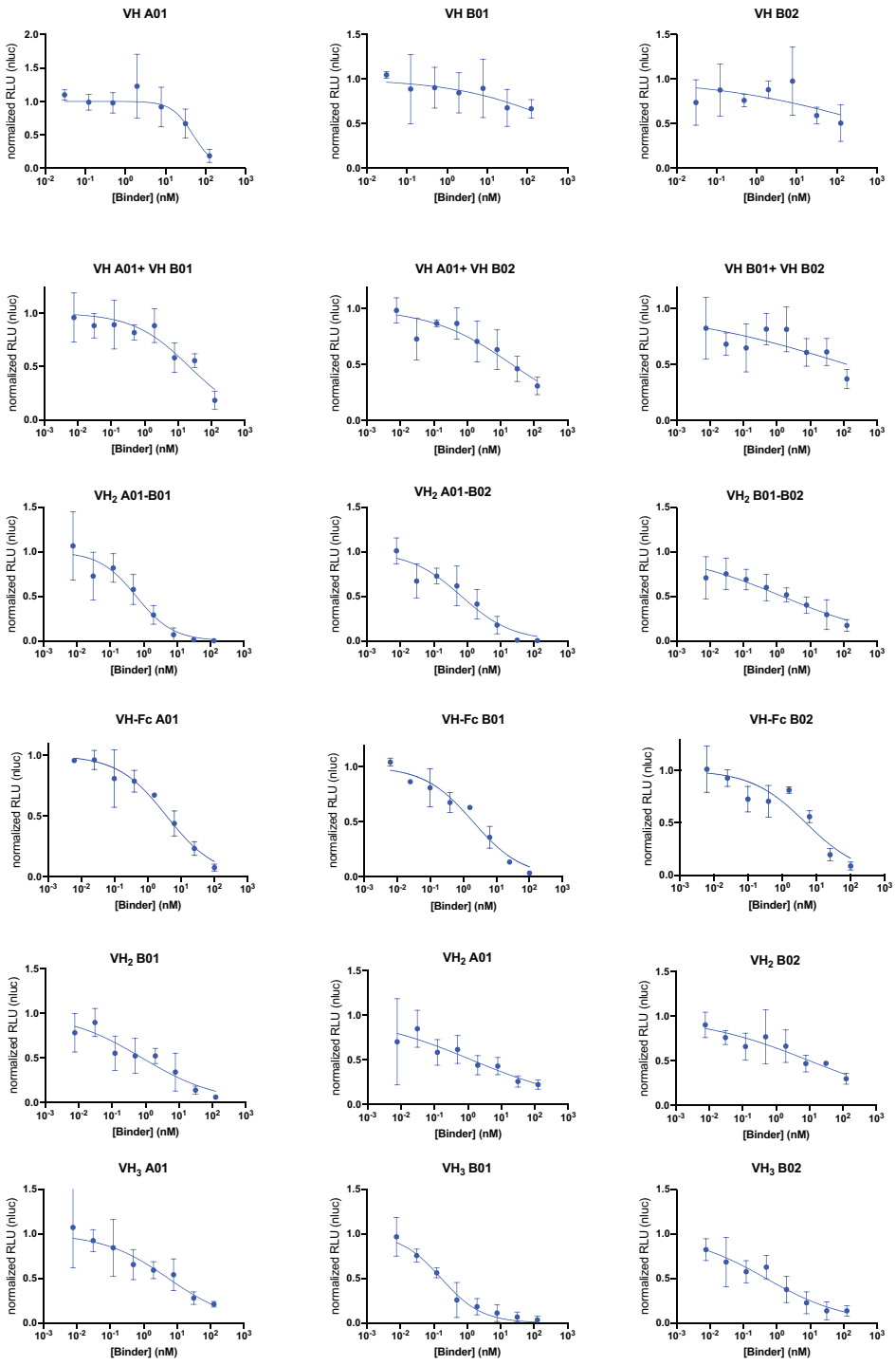


Figure 2.ED5
Pseudotyped virus neutralization assays of VH binders. Data represent average and standard deviation of two biological replicates. Data were fit to a non-linear, four-parameter variable slope regression model using Prism 8 to obtain IC50 values.

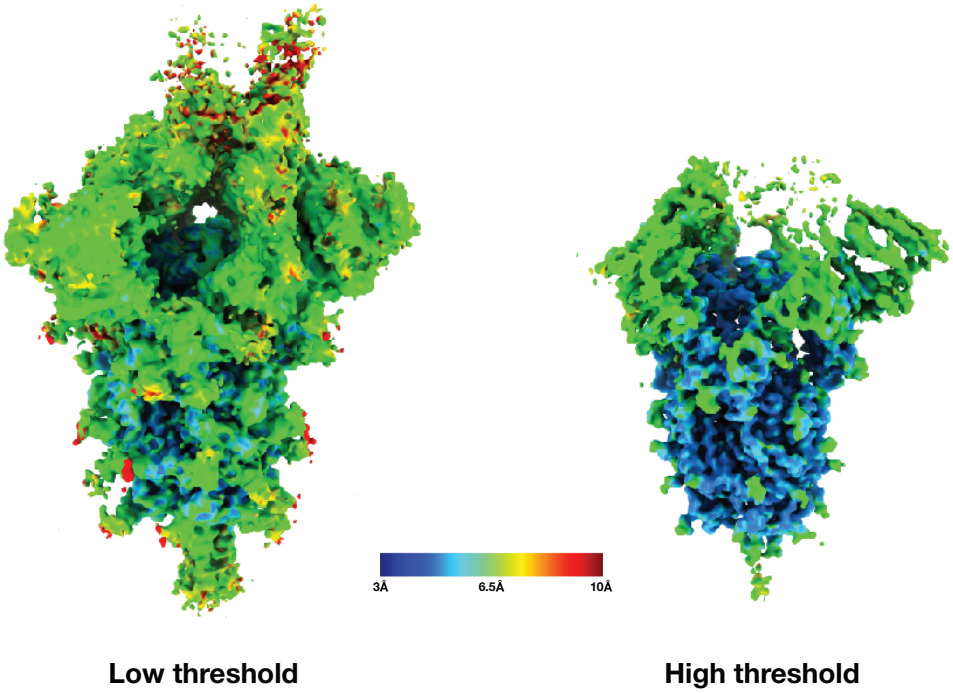


Figure 2.ED6
The SARS-CoV-2 Spike trimer + VH3 B01 cryo-EM reconstruction from non-uniform refinement in cryoSPARC at two different thresholds colored by resolution in the range from 3 Å to 10 Å. At high threshold the core S2 clearly displays high resolution features but the periphery of the molecule is closer to 6-7 Å.

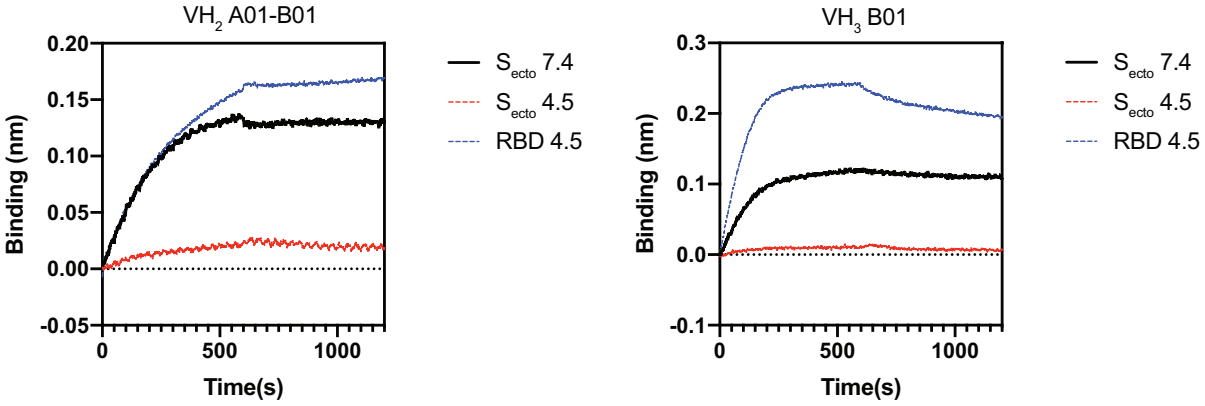


Figure 2.ED7
BLI assay of lead constructs at 25 nM to Secto and RBD at pH 7.4 or pH 4.5. Binding of VH leads is only observed with Secto at pH 7.4 and RBD at pH 4.5.

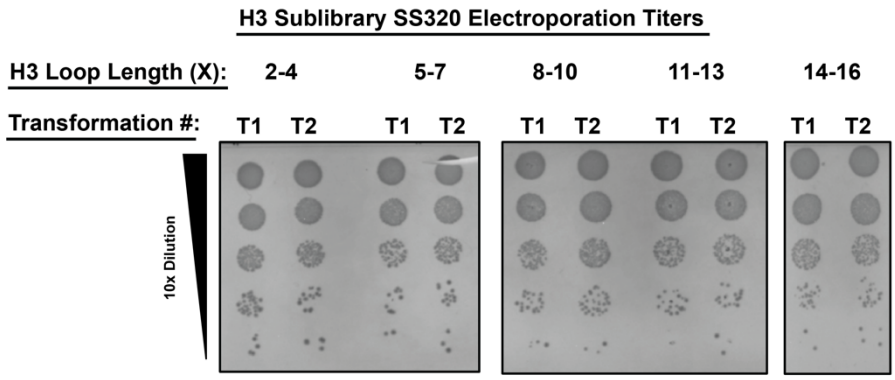
Table 2.S1: Modified VH-4D5 template sequence used for phagemid construction

Green codons represent amino acid changes (35G/39R/45E/47L/50R) in the framework. Restriction sites in each of the CDRs: AgeI in CDR H1 (blue), NcoI in CDR H2 (purple), and XhoI in CDR H3 (red).

VH-4D5 template	5'-gaggttcagctggtggagtctggcggtggcctggtgcagccaggggctcactccgtttgtcctgtgcagcttctg gcttcAACATCaccggtACTTATATCggctgggtgcgtcgcgccccgggtaagggcgaggaactggttgcacgtatc TACCCCACGccatggTATACCCGCtatgccgatagcgtcaagggccgtttcactataagcgcagacacatccaaa aacacagcctacctacaaatgaacagcttaagagctgaggacactgccgtctattattgtgctcgcTGGGGAGGGGA CGGATTCTACctcgaggactactggggtcaaggaaccctggtcaccgtctcctcg-3'

Table 2.S2: Oligonucleotides used for VH library construction denoted by their length
XXX and ZZZ are representative codons that correspond to different amino acid frequencies previously specified in **Extended Data Fig. 1**. S = C/G, W = A/T, K = G/T.

H1-5	5'–cctgtgcagcttctggcttc ZZZ ZZZ ZZZ ZZZ ZZZ ZZZ atcggctgggtgcgtcg-3'
H1-6	5'–cctgtgcagcttctggcttc ZZZ ZZZ ZZZ ZZZ ZZZ ZZZ ZZZ atcggctgggtgcgtcg-3'
H1-7	5'–cctgtgcagcttctggcttc ZZZ ZZZ ZZZ ZZZ ZZZ ZZZ ZZZ ZZZ atcggctgggtgcgtcg-3'
H2	5'–gcgaggaactggttgcacgtatc ZZZ ZZZ ZZZ ZZZ ZZZ ZZZ ZZZ ZZZ tatgccgatagcgtcaagggcc-3'
H3-4	5'– c gtc tat tat tgt gct cgc XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-5	5'– c gtc tat tat tgt gct cgc XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-6	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-7	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-8	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-9	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-10	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-11	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-12	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-13	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-14	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-15	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-16	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-17	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'
H3-18	5'– c gtc tat tat tgt gct cgc XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX XXX gSt WtK gac tac tgg ggt caa gg -3'



Sublibrary	Transformants (cfu)
2-4	1.33×10^{10}
5-7	8.33×10^{9}
8-10	1.16×10^{10}
11-13	4.99×10^{9}
14-16	1.16×10^{10}
Total:	4.98×10^{10}

Figure 2.S1: VH-phage sublibrary construction
Titers of transformed SS320 competent cells for each H3 length sublibrary are shown. The final number of transformants (cfu) was calculated as the sum from each sublibrary.

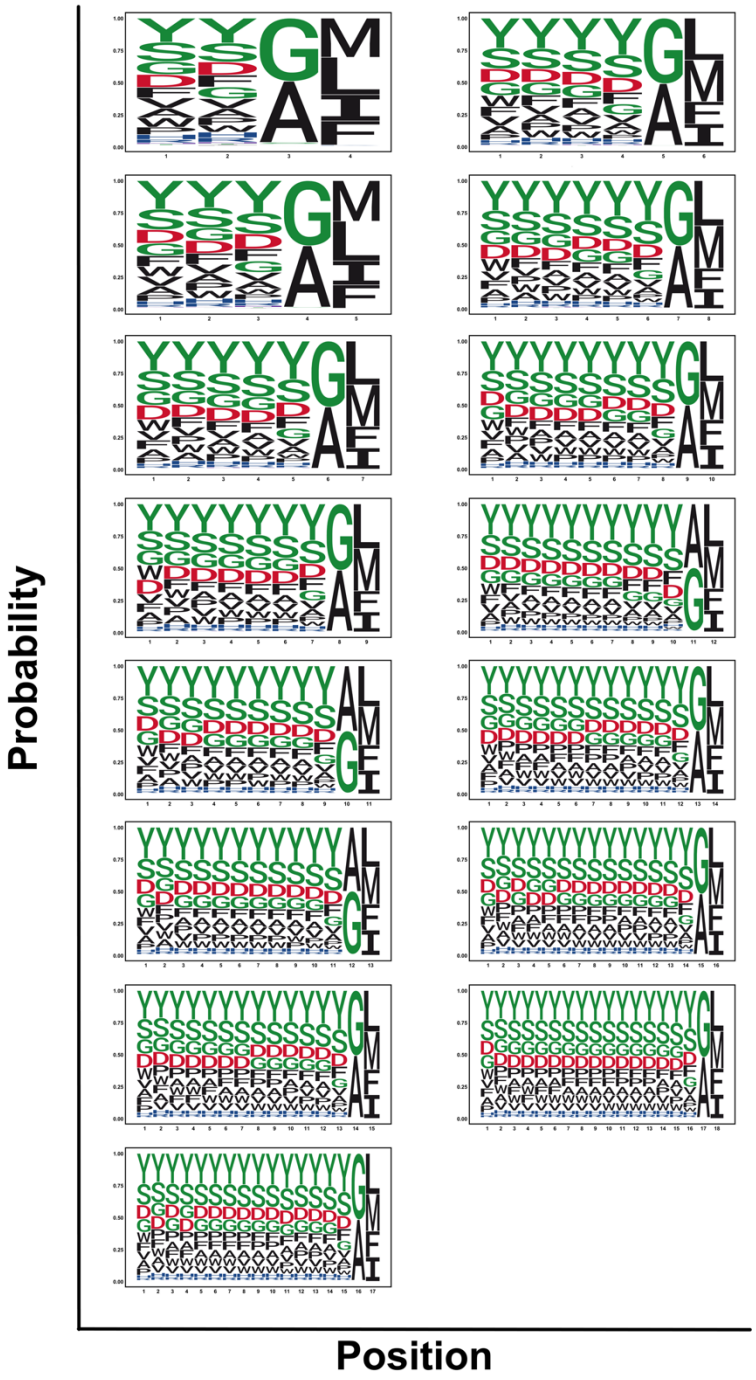


Figure 2.S2: Positional amino acid composition of all CDR H3 loops
NGS analysis shows the amino acid composition of the H3 loops (X = 2-16) for unique clones. Plots show positional frequency distribution matches designed frequencies. Position 1 refers to residue 95 (Kabat definition).

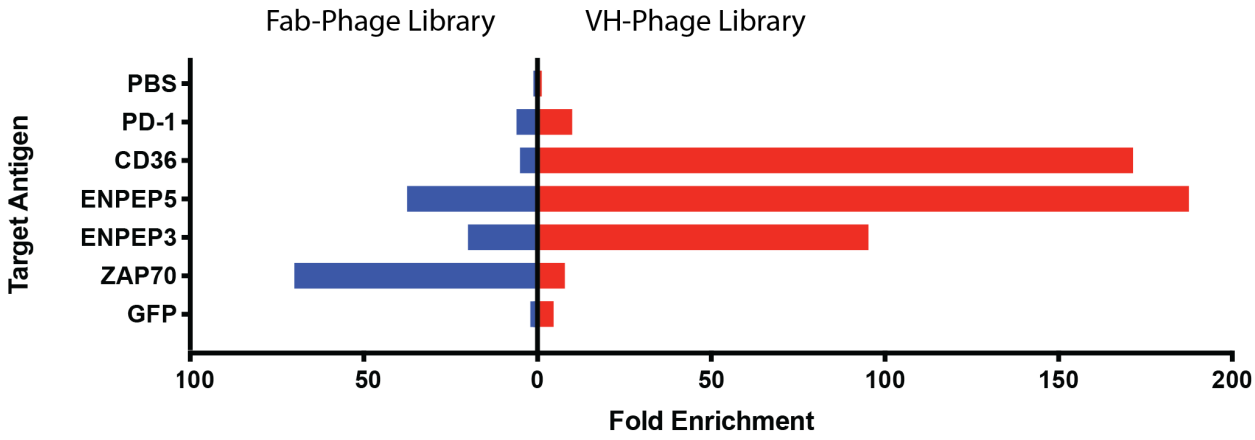


Figure 2.S3: Enrichment comparison between Fab-phage library and VH-phage library
Phage-selections were performed at the same time on a representative group of antigens. Antigens included both cytosolic and membrane proteins. Individual fold enrichment values were calculated by comparing phage titers at round three to, Fc-biotin, which was used to clear the phage pool prior to each round of selection. Phosphate buffered saline (PBS) was included as a control.

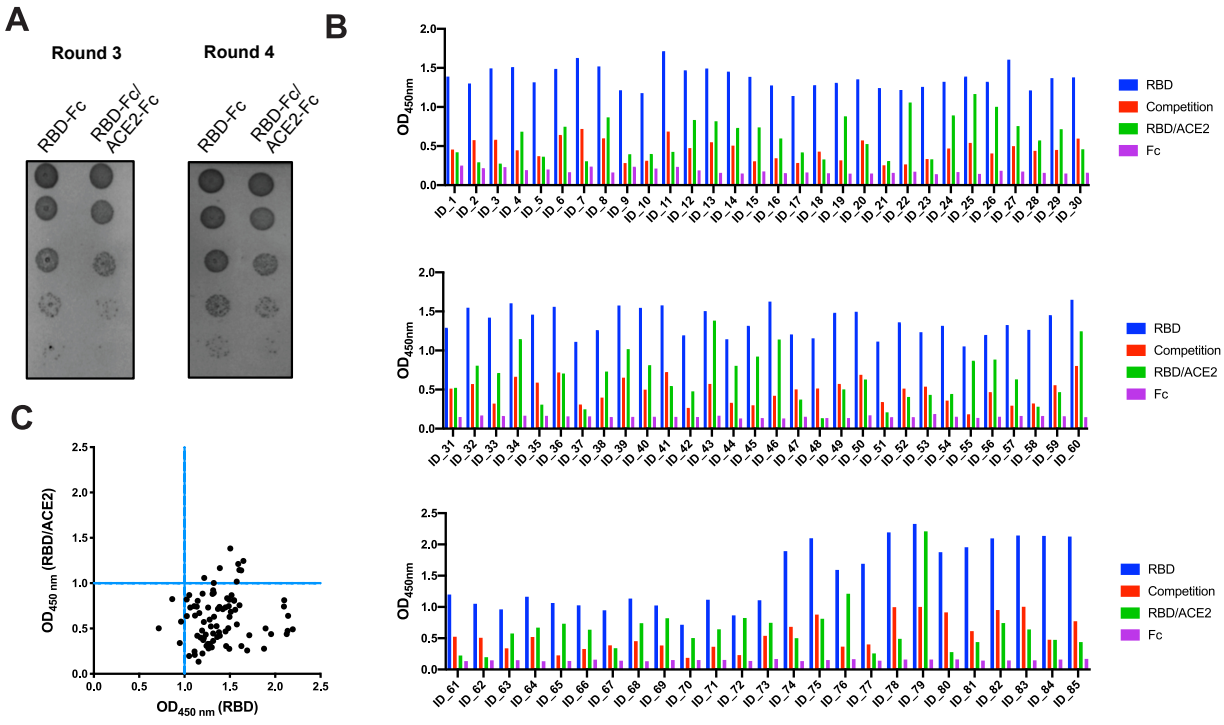


Figure 2.S4: Phage selection with VH-phage library and characterization of unique clones
(A) Titer of VH-phage after 3 or 4 rounds of selection show significant enrichment for VH-phage that bind Spike-RBD-Fc over VH-phage that bind Spike-RBD-Fc/ACE2-Fc complex. (**B**) ELISA of unique VH-phage against Spike-RBD-Fc, competition with soluble Spike-RBD-Fc, Spike-RBD-Fc/ACE2-Fc complex, or Fc antigen. (**C**) Plot of ELISA data for Spike-RBD-Fc alone vs Spike-RBD-Fc/ACE2-Fc complex shows that most of the unique VH-phage recognize the unmasked Spike-RBD-Fc antigen, suggesting they bind the same epitope as ACE2.

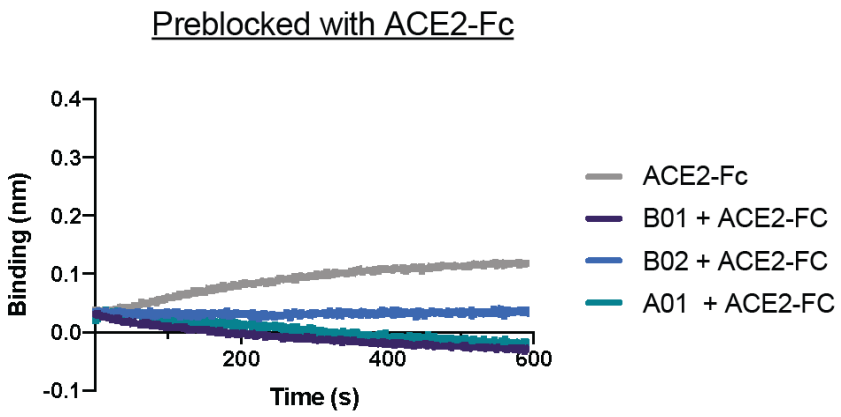


Figure 2.S5: Epitope binning of lead VH domain binders with ACE2-Fc on Spike-RBD
Sequential bio-layer interferometry (BLI) binding traces. Spike-RBD was loaded onto sensor tip and preblocked with ACE2-Fc, followed by adding either VH A01, VH B01, or VH B02. These association data indicate that ACE2-Fc prevents binding of either of the three VH domains, suggesting they share the same binding epitope on Spike-RBD.

Table 2.S3: ELISA EC50 of VH Binders

VH ID	EC50 (nM)
VH A01	3.55 ± 0.67
VH B01	12.0 ± 1.6
VH B02	14.5 ± 4.3
VH-Fc A01	0.0738 ± 0.011
VH-Fc B01	0.265 ± 0.042
VH-Fc B02	0.0415 ± 0.002
VH^2 A01-B01	0.139 ± 0.027
VH^2 A01-B02	0.208 ± 0.014
VH^2 B01-B02	0.295 ± 0.049
VH^2 A01	0.280 ± 0.067
VH^2 B01	0.230 ± 0.053
VH^2 B02	0.217 ± 0.001
VH^3 A01	0.323 ± 0.023
VH^3 B01	0.135 ± 0.011
VH^3 B02	0.328 ± 0.032

Table 2.S4: BLI binding affinity of VH binders to Spike-RBD

VH ID	Spike-RBD K_D (nM)			
	VH	VH-Fc	VH_2	VH_3
A01	23	8.40	1.89	4.00
B01	46	0.12	0.19	0.11
B02	113	0.19	0.46	5.18
A01-B01			0.10	
A01-B02			0.48	
B01-B02			0.23	

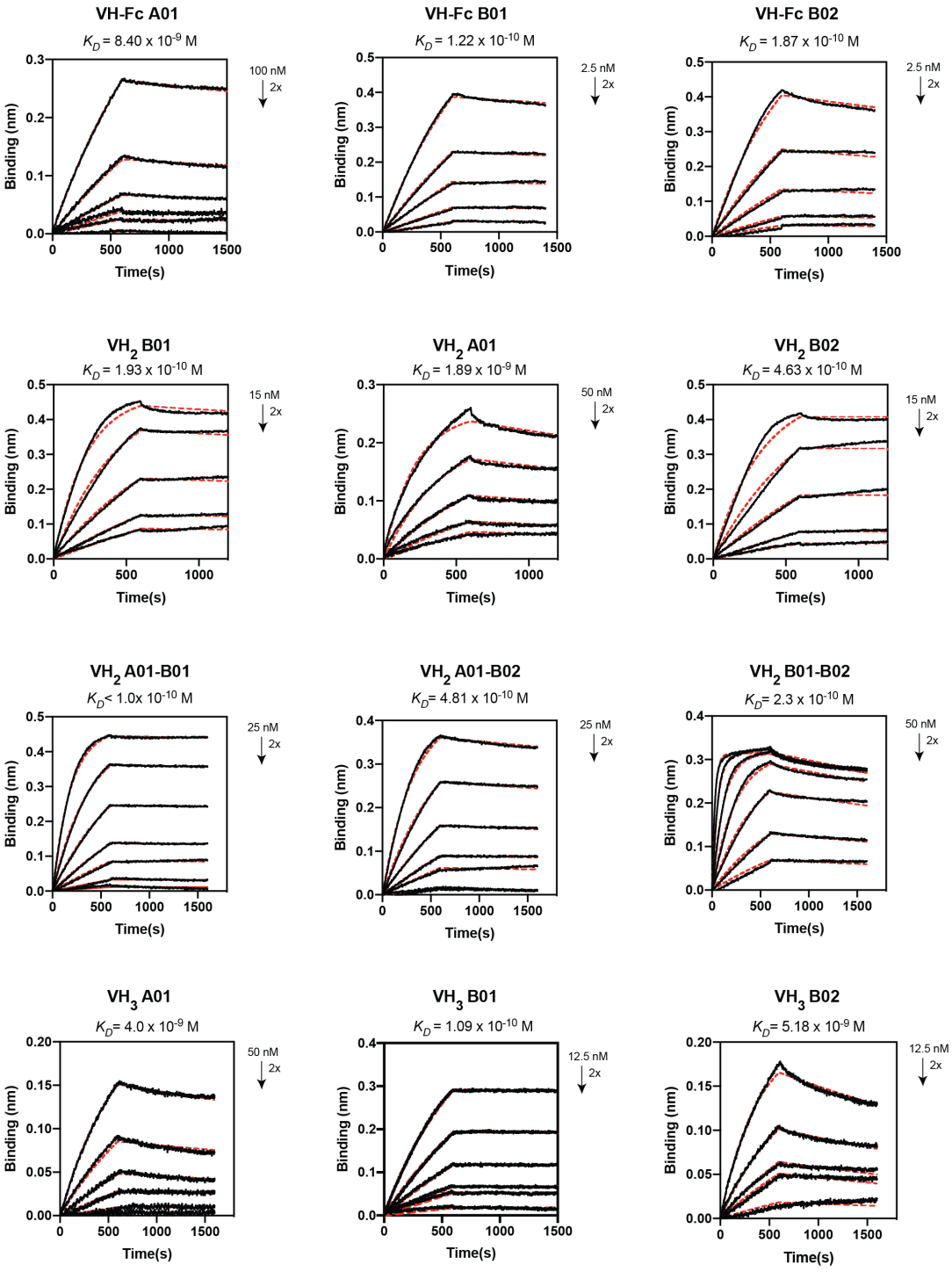


Figure 2.S6: Multipoint Biolayer Interferometry (BLI) of VH Binders to Spike-RBD
Multipoint BLI experiments for each VH binder. The data were fit to a 1:1 binding model using the Octet ForteBio software to determine the binding affinity for the measured interaction.

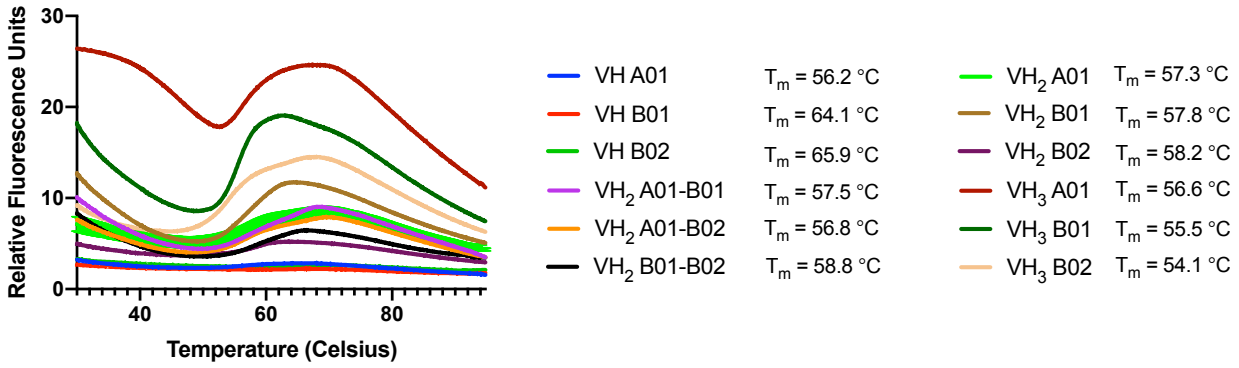


Figure 2.S 7: Differential Scanning Fluorimetry (DSF) of VH Binders
The melting temperature (T_m) of each VH binder was determined by differential scanning fluorimetry. Data and T_m presented are an average of two replicates.

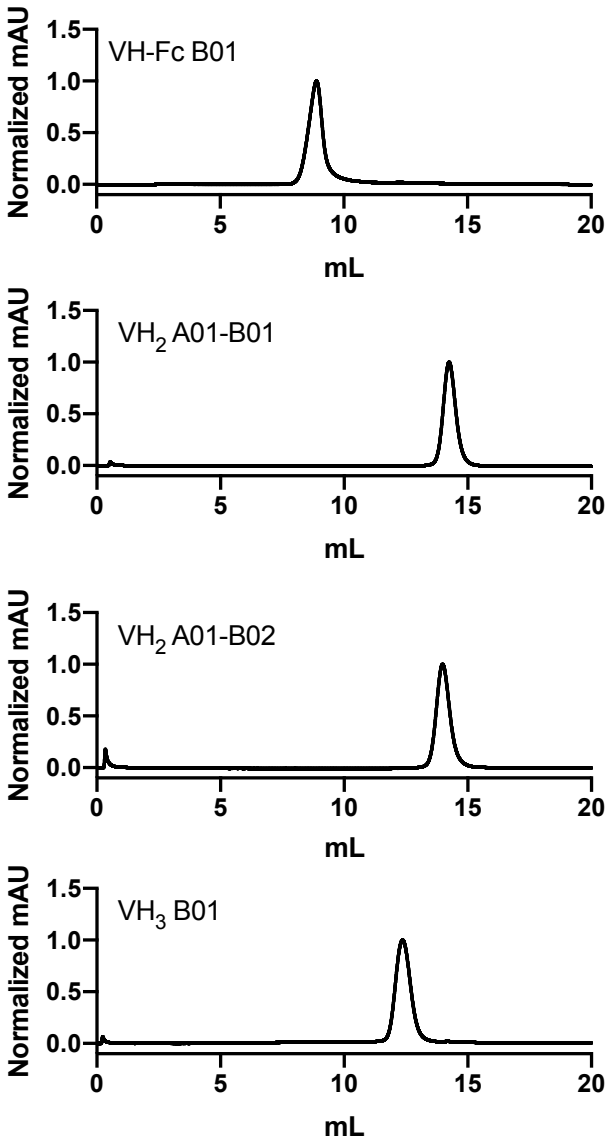


Figure 2.S8: Size Exclusion Chromatography of VH Binders
Size exclusion chromatography (SEC) traces of VH-Fc B01, VH_2 A01-B01, VH_2 A01-B02, and VH_3 B01 on a Superdex 200 Increase 10/300 GL column.

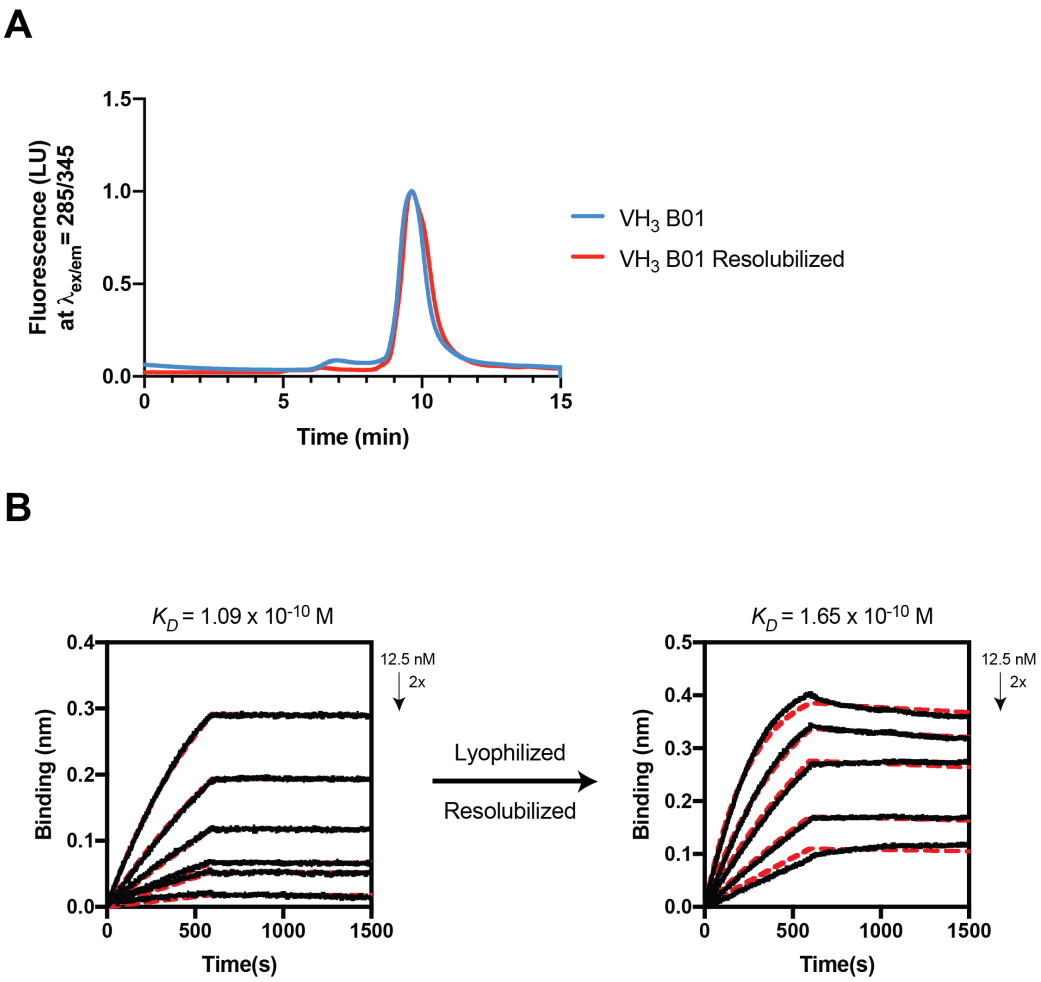


Figure 2.S9: Lyophilization and reconstitution of VH_3 B01
(A) SEC trace of VH_3 B01 on an AdvanceBio 300 Å column before and after lyophilization shows that the elution profile of VH_3 B01 is unchanged. (**B**) Multipoint BLI of VH_3 B01 binding to Spike-RBD before and after lyophilization and resolubilization show that the binding affinity of VH_3 B01 is largely unchanged.

Table 2.S5: Pseudotyped virus neutralization IC50

VH ID	Pseudotyped virus			
	IC_{50} nM (ng/mL)			
	VH	VH-Fc	VH_2	VH_3
A01	55.1 ± 3.3 (763 ± 46)	2.67 ± 0.85 (209 ± 66)	1.75 ± 0.77 (50.7 ± 22.3)	4.76 ± 2.28 (210 ± 101)
B01	218 ± 70 (3032 ± 978)	1.86 ± 0.24 (145 ± 19)	1.55 ± 1.45 (45.0 ± 42.2)	0.156 ± 0.048 (6.90 ± 2.12)
B02	156 ± 28 (2065 ± 378)	5.93 ± 1.95 (462 ± 152)	14.0 ± 16.8 (388 ± 466)	0.396 ± 0.176 (16.7 ± 7.46)
A01 + B01	28.9 ± 15.5 (400 ± 215)			
A01 + B02	35.3 ± 34.7 (491 ± 482)			
B01 + B02	423 ± 193 (5608 ± 2554)			
A01-B01			0.74 ± 0.28 (21.4 ± 8.1)	
A01-B02			1.08 ± 0.67 (30.6 ± 18.9)	
B01-B02			2.90 ± 0.79 (82.4 ± 22.4)	

Table 2.S6: Live SARS-CoV-2 Neutralization IC50

VH ID	SARS-CoV-2 IC50 nM (ng/mL)
VH_2 A01-B01	12.0 ± 3.0 (347 ± 87)
VH_2 A01-B02	26.2 ± 6.5 (743 ± 184)
VH-Fc B01	33.5 ± 8.3 (2613 ± 647)
VH_3 B01	3.98 ± 1.56 (176 ± 69)

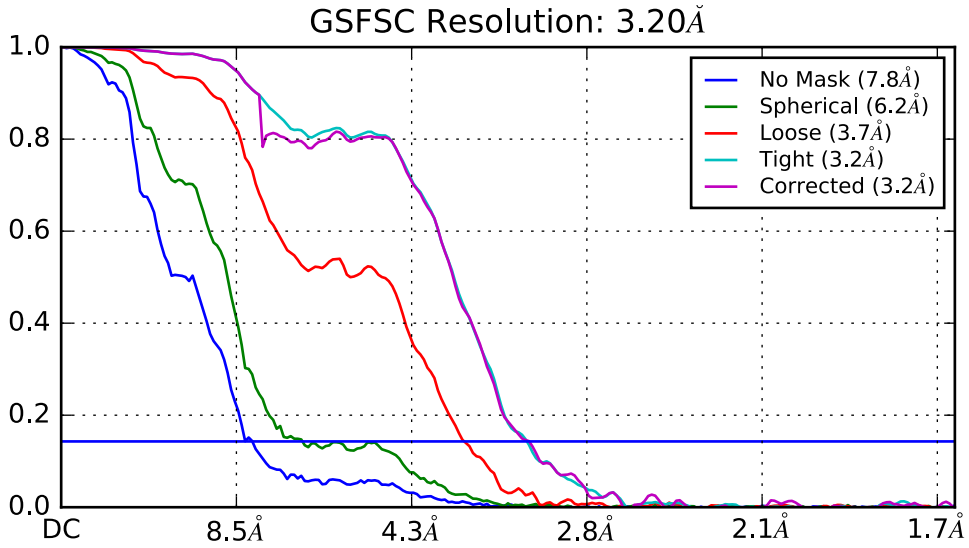


Figure 2.S10: Gold standard FSC curves from final iteration of non-uniform refinement in cryoSPARC for masked and unmasked reconstructions.

Table 2.S7: Cryo-EM data collection, refinement and validation statistics

	(EMDB ID: EMD-22514) (PDB ID: 7JWB)
Data collection and processing	
Magnification	105,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	78
Defocus range (μm)	0.7-2.8
Pixel size (Å)	0.834
Symmetry imposed	C1
Initial particle images (no.)	317000
Final particle images (no.)	20831
Map resolution (Å)	3.2
FSC threshold	0.143
Map resolution range (Å)	3.2-10
Refinement	
Initial model used (PDB code)	6X2B, 4G80
Model resolution (Å)	7.05
FSC threshold	0.5
Model resolution range (Å)	4-8
Map sharpening *B* factor (Å^2)	50
Model composition	
Non-hydrogen atoms	25657
Protein residues	3289
Ligands	0
B factors (Å^2)	
Protein	N/A
Ligand	N/A
R.m.s. deviations	
Bond lengths (Å)	0.019
Bond angles (°)	1.703
Validation	
MolProbity score	0.72
Clashscore	0.67
Poor rotamers (%)	0
Ramachandran plot	
Favored (%)	98.19
Allowed (%)	1.65
Disallowed (%)	0.16

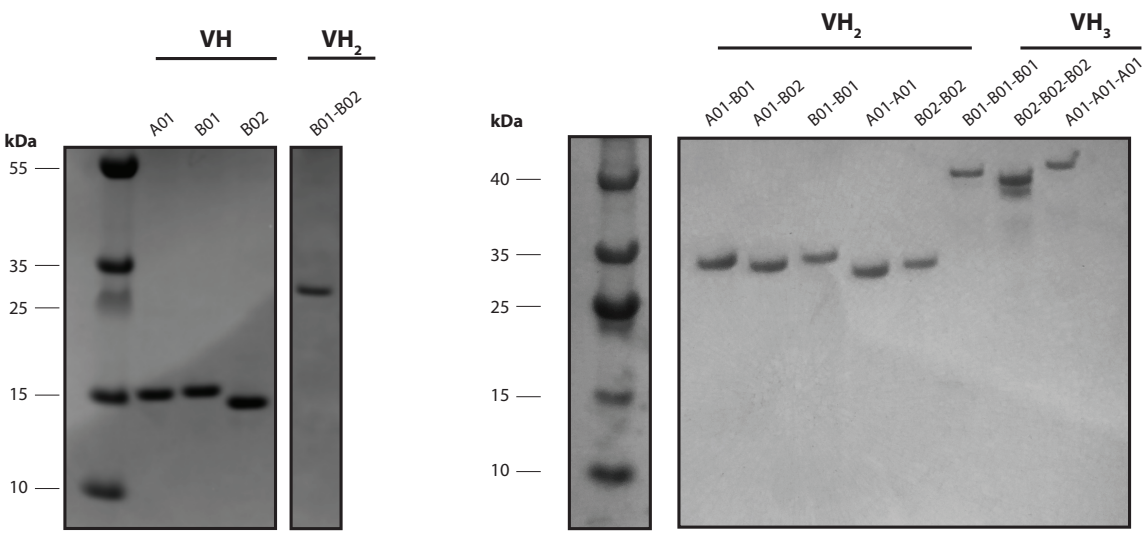


Figure 2.S11: SDS-PAGE gel image of Protein A purified VH constructs
SDS-PAGE analysis of expressed VH constructs. The two separate gel images are representative for the Protein A purified VH constructs used in this study.

2.8 References

Main Text References

1. Yan, R. *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).
2. Abraham, J. Passive antibody therapy in COVID-19. *Nat. Rev. Immunol.* **20**, 401–403 (2020).
3. Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (2020).
4. Pinto, D. *et al.* Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020).
5. Rogers, T. F. *et al.* Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* **7520**, eabc7520 (2020).
6. Cao, Y. *et al.* Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* **182**, 73–84 (2020).
7. Shi, R. *et al.* A human neutralizing antibody targets the receptor binding site of SARS-CoV-2. *Nature* **584**, 120–124 (2020).
8. Miersch, S. *et al.* Scalable high throughput selection from phage-displayed synthetic antibody libraries. *J. Vis. Exp.* 1–15 (2015). doi:10.3791/51492
9. Nguyen, V. K., Hamers, R., Wyns, L. & Muyldermans, S. Camel heavy-chain antibodies: Diverse germline V(H)H and specific mechanisms enlarge the antigen-binding repertoire. *EMBO J.* **19**, 921–930 (2000).
10. Holt, L. J., Herring, C., Jespers, L. S., Woolven, B. P. & Tomlinson, I. M. Domain antibodies: Proteins for therapy. *Trends Biotechnol.* **21**, 484–490 (2003).

11. Ingram, J. R., Schmidt, F. I. & Ploegh, H. L. Exploiting Nanobodies' Singular Traits. *Annu. Rev. Immunol.* **36**, (2018).

12. Hamer-Casterman Atarchouch,T, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1998).

13. De Genst, E. *et al.* Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4586–4591 (2006).

14. Arbabi-Ghahroudi, M., Tanha, J. & MacKenzie, R. Prokaryotic expression of antibodies. *Cancer Metastasis Rev.* **24**, 501–519 (2005).

15. Zimmermann, I. *et al.* Synthetic single domain antibodies for the conformational trapping of membrane proteins. *Elife* **7**, 1–32 (2018).

16. McMahon, C. *et al.* Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat. Struct. Mol. Biol.* **25**, 289–296 (2018).

17. Saerens, D., Ghassabeh, G. H. & Muyldermans, S. Single-domain antibodies as building blocks for novel therapeutics. *Curr. Opin. Pharmacol.* **8**, 600–608 (2008).

18. Nilvebrant, J., Tessier, P. & Sidhu, S. Engineered Autonomous Human Variable Domains. *Curr. Pharm. Des.* **22**, 6527–6537 (2016).

19. Rouet, R., Dudgeon, K., Christie, M., Langley, D. & Christ, D. Fully human VH single domains that rival the stability and cleft recognition of camelid antibodies. *J. Biol. Chem.* **290**, 11905–11917 (2015).

20. Sun, Z. *et al.* Potent neutralization of SARS-CoV-2 by human antibody heavy-chain variable domains isolated from a large library with a new stable scaffold. *MAbs* **12**, 1778435 (2020).

21. Cao, Y. *et al.* Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-

Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell* **182**, 73–84 (2020).

22. Huo, J. *et al.* Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. *Nat. Struct. Mol. Biol.* **27**, 846–854 (2020).

23. Walter, J. D. *et al.* Sybodies targeting the SARS-CoV-2 receptor-binding domain. *BioRxiv* (2020). doi:10.1101/2020.04.16.045419

24. Palomo, C. *et al.* Trivalency of a nanobody specific for the human respiratory syncytial virus fusion glycoprotein drastically enhances virus neutralization and impacts escape mutant selection. *Antimicrob. Agents Chemother.* **60**, 6498–6509 (2016).

25. Carter, P. *et al.* Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4285–4289 (1992).

26. Harmsen, M. M. *et al.* Llama heavy-chain V regions consist of at least four distinct subfamilies revealing novel sequence features. *Mol. Immunol.* **37**, 579–590 (2000).

27. Ma, X., Barthelemy, P. A., Rouge, L., Wiesmann, C. & Sidhu, S. S. Design of synthetic autonomous VH domain libraries and structural analysis of a VH domain bound to vascular endothelial growth factor. *J. Mol. Biol.* **425**, 2247–2259 (2013).

28. Dudgeon, K. *et al.* General strategy for the generation of human antibody variable domains with increased aggregation resistance. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10879–10884 (2012).

29. Birtalan, S. *et al.* The Intrinsic Contributions of Tyrosine, Serine, Glycine and Arginine to the Affinity and Specificity of Antibodies. *J. Mol. Biol.* **377**, 1518–1528 (2008).

30. Hornsby, M. *et al.* A High Through-put Platform for Recombinant Antibodies to Folded Proteins. *Mol. Cell. Proteomics* **14**, 2833–2847 (2015).

31. Martinko, A. J. *et al.* Targeting RAS-driven human cancer cells with antibodies to upregulated and essential cell-surface proteins. *Elife* **7**, e31098 (2018).

32. Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H., Gottesman, K. *Sequences of proteins of immunological interest*. *National Institutes of Health, Bethesda, MD, U.S.A.* (1987). doi:10.1016/0003-2697(84)90805-4

33. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).

34. Lui, I. *et al.* Trimeric SARS-CoV-2 Spike interacts with dimeric ACE2 with limited intra-Spike avidity. *BioRxiv* (2020). doi:10.1101/2020.05.21.109157

35. Neuman, B. W. *et al.* Supramolecular Architecture of Severe Acute Respiratory Syndrome Coronavirus Revealed by Electron Cryomicroscopy. *J. Virol.* **80**, 7918–7928 (2006).

36. Byrnes, J. R. *et al.* Competitive SARS-CoV-2 Serology Reveals Most Antibodies Targeting the Spike Receptor-Binding Domain Compete for ACE2 Binding. *mSphere* **5**, e00802-20 (2020).

37. Crawford, K. H. D. *et al.* Protocol and reagents for pseudotyping lentiviral particles with SARS-CoV-2 spike protein for neutralization assays. *Viruses* **12**, 513 (2020).

38. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **180**, 281–292 (2020).

39. Zhou, T. *et al.* A pH-dependent switch mediates conformational masking of SARS-CoV-2 spike. *BioRxiv* (2020). doi:10.1101/2020.07.04.187989

40. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).

41. Dong, J. *et al.* Development of multi-specific humanized llama antibodies blocking SARS-

CoV-2/ACE2 interaction with high affinity and avidity. *Emerg. Microbes Infect.* **9**, 1034–1036 (2020).

42. Binz, H. K., Amstutz, P. & Plückthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* **23**, 1257–1268 (2005).

43. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503 (2003).

44. Binz, H. K. *et al.* Design and characterization of MP0250, a tri-specific anti-HGF/anti-VEGF DARPin® drug candidate. *MAbs* **9**, 1262–1269 (2017).

45. Conrath, K. E., Lauwereys, M., Wyns, L. & Muyldermans, S. Camel Single-domain Antibodies as Modular Building Units in Bispecific and Bivalent Antibody Constructs. *J. Biol. Chem.* **276**, 7346–7350 (2001).

46. Strauch, E. M. *et al.* Computational design of trimeric influenza-neutralizing proteins targeting the hemagglutinin receptor binding site. *Nat. Biotechnol.* **35**, 667–671 (2017).

47. Galimidi, R. P. *et al.* Intra-spike crosslinking overcomes antibody evasion by HIV-1. *Cell* **160**, 433–446 (2015).

Methods References:

48. Lee, C. V. *et al.* High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold. *J. Mol. Biol.* **340**, 1073–1093 (2004).

49. Chen, G. & Sidhu, S. S. Design and generation of synthetic antibody libraries for phage display. *Methods Mol. Biol.* **1131**, 113–131 (2014).

50. Wang, B. *et al.* Bivalent binding of a fully human IgG to the SARS-CoV-2 spike proteins reveals mechanisms of potent neutralization. *BioRxiv* (2020).

doi:10.1101/2020.07.14.203414

51. Zheng, S. Q. *et al.* MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

52. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

53. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: Adaptive regularization improves single particle cryo-EM reconstruction. *BioRxiv* (2019). doi:10.1101/2019.12.15.877092

54. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

55. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).

56. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

Chapter 3

Single domains antibodies provide small-molecule switches for CAR T-Cell Therapy

3.1 Abstract

Minimizing the frequent toxic side-effects associated with immunotherapies is necessary to broaden their application to solid tumors. While a host of chemically inducible and chemical genetic off-switches have been applied to these therapies, there is no routine method for generating small intracellular switches to multiple clinically validated small-molecules. Here we utilize a phage display strategy to generate multiple single-domain antibodies (sdAbs) capable of recognizing several E3-ligase-small molecule complexes, thus turning the ligand into dimerizer molecules. Simultaneously, we find sdAbs that can do the opposite. Recognize E3-ligase small molecule binding pocket thus turning the ligand into disrupter molecule. We further engineer one of these dimerizers and demonstrate its ability to degrade chimeric antigen receptors (CARs) in primary human T-cells.

3.2 Introduction

T cells engineered to express CARs have been transformative for patients with relapsed and refractory B-cell malignancies. The excitement surrounding this modality is warranted by a durable and a high overall response, however the substantial clinical benefit of these therapies is frequently hindered by life-threatening side effects, such as cytokine release syndrome, neurotoxicity, and B cell aplasia [1-4]. While the most severe of these adverse events are manageable

in a clinical setting, safety remains a barrier for CAR-T cells targeting antigens on solid tumors as the likelihood of on-target off-tumor toxicities is higher. Therefore, the ability to switch or control CAR-T cell activity on command can hasten the development of these immunotherapies and expand their clinical benefit.

A substantial number of strategies to chemically control CAR-T cell response have been reported outside of clinical studies. These include strategies to induce apoptosis, either by small-molecule or antibodies targeting CAR-T cell antigens [5,6]. Some strategies offer a split system of dissociated recognition domains and signaling domains, while others have used multi-specific antibodies or proteins to conditionally potentiate CAR-T cells [7,8]. Most recently, chemical genetic control systems have been applied to regulate the degradation of CARs. These ligand-induced degradation systems include proteolysis targeting chimeras (PROTACs), destabilized domains or dTags, and the CRBN/IMiD-binding degron of IKZF1-IKZF3 [9,10]. While these systems demonstrate varied level of success, there is a need for multiple chemical genetic control system that are controlled by clinically validated small molecules that are derived from human scaffolds.

Using these parameters, we demonstrate an approach for turning multiple ligands of E3 ubiquitin ligases into dimerizer molecules capable of inducing proximity of the E3 ligase and CAR (**Fig 3.1a**). To do this, we use phage display to generate highly specific sdAbs to the ligand-E3 complex (**Fig 3.1b**). Serendipitously, we also generate sdAbs to E3 ligand binding site. Thus, turning the ligand into a disrupter molecule capable of disrupting the proximity between an E3 ligase and sdAb (**Fig 3.1c**). We further engineer one of these dimerizers into an off-switch capable of inducibly degrading CARs in primary T-cells.

3.3 Results

Generation of chemical-epitope-selective sdAbs

We first selected two sets of E3-ligase ligands that are well-tolerated in humans and have been previously utilized to degrade non-substrate proteins. The first set consists of thalidomide and its immunomodulatory drug (IMiD) derivatives that bind the E3 ligase cereblon (CRBN) (**Fig 3.2a-b**). The IMiDs are approved by the U.S. Food and Drug Administration (FDA) as proven therapies for multiple myeloma and other indications. Their mechanism of action involves ubiquitination and subsequent degradation of neo-substrates [11-14]. Additionally, they have been used in hetero-bifunctional degraders to degrade non-substrate proteins. Ideally, a large portion of the small molecule needs to remain solvent exposed when bound to generate highly specific antibodies for the chemical epitope [15]. Conversely, IMiDs have a very small solvent exposed area, however, IMiD binding to CRBN drastically alters substrate specificity, presumably by a conformational change [11]. We reasoned that this structural difference between bound and unbound could be leveraged to generate highly specific sdAbs. The second set of molecules consists of the SMAC mimetics that bind the E3 ligase cellular inhibitor of apoptosis protein 1 (cIAP1) (**Fig 3.2 e-f**). These molecules are based on a conserved tetra-peptide IAP-binding motif and have good solvent accessibility when bound to cIAP1. Additionally, they have been utilized as hetero-bifunctional degraders. While the SMAC mimetics are not FDA approved due to lack of efficacy as a cIAP1 antagonist, they were tolerated in clinical trials [16].

To identify unique chemical-epitope-selective antibodies, we expressed and purified the ligand binding domain and the BIR3 domain of cIAP. Utilizing a high diversity sdAb library we performed phage-display selections on the protein-ligand complex [15]. To specifically enrich for epitope-specific binders to the complex we first cleared the library with E3-ligase alone as was

previously described [17]. By rounds 3 and 4, significant enrichments for phage that bound the protein-ligand but not the protein alone were observed. Single clones were isolated and characterized for their ability to bind the E3-Ligase-ligand complex by phage enzyme-linked immunosorbent assays. In total, 53 unique clones were identified. Thirteen sdAbs passed the threshold of acceptable two-fold specificity to the SMAC-cIAP1 complex. Only one unique clone was found for lenalidomide-CRBN selection, however it showed a very high six-fold specificity for the lenalidomide-CRBN complex (**Supplemental Figure 3.1**). The dominance of this clone suggests that the selection was bottle-necked by round 4, and that potentially more unique clones could be identified in the outputs of earlier rounds.

Characterization of dimerizers and disrupters

To further profile the best sdAbs we expressed and characterized the kinetics of binding. The sdAb (VH107) targeting lenalidomide-CRBN bound with moderate affinity (K_D=350nM) and had no detectable binding to CRBN in the absence of lenalidomide (**Fig 3.2c**) even in concentrations up to 5000nM VH107 (data not shown). Thus, demonstrating VH107 as a highly specific dimerizer. We next tested binding as a function of lenalidomide concentration given saturating amounts of VH107 and observed an EC_{50} of 167nM (**Fig 3.2d**). This observation was encouraging as a lenalidomide concentration of hundreds of nanomolar can be achieved unbound in plasma using a single oral dose of 30mg/kg in mice for 15-20 hours, suggesting VH107 could be utilized as a switch in mouse models [9]. Notably, this EC_{50} differs greatly from the reported affinities of Lenalidomide for CRBN (~3uM) [18]. This suggested cooperative binding indicates that VH107 could be stabilizing an IMiD binding conformation of CRBN and not directly interacting with the small molecule. We reasoned that if this were the case VH107 would recognize all three

IMiDs even though VH107 was identified in a selection with lenalidomide. As predicted, we observed comparable binding between VH107 and all three IMiD-CRBN complexes (**Fig 3.2i**) suggesting VH107 binds an IMiD binding conformation of CRBN.

Next, we tested sdAbs isolated from SMAC-cIAP1selections. While several clones bound with moderate specificity and affinity to AT406-cIAP1 over cIAP1 alone, we were surprised to find several sdAbs that bound cIAP1 but not SMAC-cIAP1 complex (**Supplemental Figure 3.2**). The highest affinity of these sdAbs (VH108) had a moderate binding affinity of 144nM for cIAP1 (**Fig 3.2g**). This specificity suggests that VH108 is binding in the small molecule binding pocket of cIAP1. A dose response of binding at varying concentrations of birinapant show an IC_{50} of 2nM, which tracks well with the single-digit nanomolar affinity reported for birinapant (**Fig 3.2h**). These data suggest that birinapant can be used as a chemical disrupter to break apart VH108-cIAP1.

To develop these dimerizers and disrupters as control switches for CAR T cells we next wanted to test their ability to bind reversibly in mammalian cells. To visualize dimerization or disruption of sdAb-E3 ligase partners we utilized a fluorophore phase transition-based assay called separation of phases-based protein interaction reporter (SPPIER) [19,20]. The system consists of three domains, the sdAB or E3 ligase of interest, enhanced GFP (EGFP), and a homo-oligomeric tag (HOTag) (**Fig 3.3a**). Cells were transfected with both sdAB and E3 ligase constructs and treated with 2.5μM IMiD or birinapant after 24hrs. Immediately after adding drug cells were imaged 60x 30-second intervals apart. Gratifyingly, the VH107-CRBN construct phase separated and formed bright EGFP droplets (**Fig 3.3b**). Validating VH107 functionality in an intracellular environment. Unfortunately, VH108-cIAP1 did not show inducible disruption, and no EGFP droplets were observed after transfection. Upon increasing the expression levels of the VH108 construct

nonspecific aggregation of EGFP was observed (**Supplemental Fig 3.3b**). This suggests that VH108 is misfolding in the intracellular environment and nonfunctional in cells.

Engineering of VH107 degron for primary CAR-T cells

We next sought to use VH107 to regulate the surface expression of a CAR. We utilized the CAR of the clinically approved tisagenlicleucel, which is approved for B cell-precursor acute lymphoblastic leukemia. This construct comprised of a single polypeptide with a humanized anti-CD19 single-chain variable fragment (scFv) linked to a CD8 transmembrane domain, a 4-1BB costimulatory domain, and a CD3ζ signaling domain. To make VH107 into an off-switch for the CAR the sdAb was fused to the C-term of the CD3ζ to recruit endogenous CRBN and associated degradation machinery to the CAR. Two additional modifications were made to the CAR, a thosea asigna virus 2A (T2A) self-cleavage site and EGFP were appended to the C-term of VH107 so that total CAR expression could be monitored, and an extracellular c-Myc-tag was appended to the N-term of the scFv to measure cell-surface levels of CAR. Primary human T cells were then transduced with lentivirus containing either the CD19-VH107 CAR, CD19 CAR, or a CD19-IKZF3 CAR as a positive control. The CD19-IKZF3 CAR contains a miniaturized IKZF3 zinc finger motif and has been reported to degrade CARs upon addition of lenalidomide through recruitment of CRBN [9]. Crucially, upon a dose titration of lenolidomide to CD19-VH107 CAR surface levels of the CAR decrease in a dose-dependent way, with the highest dosage completely reducing expression levels to background as compared to untransduced cells (UTD) (**Fig 3.4a**). This result demonstrates that CAR expression levels can be tuned with varying amounts of lenalidomide. Notably, the magnitude of degradation by CD19-VH107 is comparable to the positive control IKZF3-degron at the maximal dose of lenalidomide (**Fig 3.4c**).

Overall transduction efficiency was comparable between unmodified CAR and both degrons, however, the surface levels of CAR were several-fold lower in both degron constructs as compared to unmodified (**Fig 3.4b**). This result could suggest several possibilities: a basal level of degradation by both CAR-degrons; c-terminal fusions destabilize the CAR; or instability of VH107, caused by the absence of the intrachain disulfide bond in the reducing cytosol, could contribute to this lower surface expression. To assess the stability of VH107 in absence of the disulfide, recombinant VH107 was resuspended in a mild reducing condition through the addition of β-mercaptoethanol (BME) and subjected to differential scanning fluorimetry (DSF). A ~11°C decrease in melting temperature (T_m) was observed (**Supplemental Table 3.1**) suggesting instability in the cytosol. Simultaneously, we made a disulfide scrubbed mutant (C22A, C95A) of VH107 and attempted expression in *E.coli* and as an Aga2 surface fusion in *Saccharomyces cerevisiae*, however no expression was detected during purification or via flow cytometry (data not shown).

We reasoned that an affinity maturation campaign using yeast surface display and fluorescence activated cell sorting (FACS) could enable the discovery of a variant with improved stability and affinity. To create mutations across the entire VH domain we utilized autonomous hypermutation yeast surface display (AHEAD), which enables continuous hypermutation of the gene of interest at a rate 100,000-fold higher than yeast's genomic mutation rate [21]. After successful integration of VH107 into the AHEAD yeast strain, 8 rounds of FACS were performed incrementally decreasing CRBN concentration (with 5uM lenalidomide) until a final of 250pM at R8. Gratifyingly, output yeast showed stronger binding and expression by flow cytometry compared to the parental clone (**Fig5a**). Amplicon sequencing was performed to track clone abundance round to round and showed that R8 clones had mutations in both the CDRs and the

scaffold, suggesting possible stabilizing mutations (**Fig 3.5b-c**). Indeed, recombinant expression of the R8 winner (VH107mR94) showed an increase in yield as compared to the parental clone. Kinetic analysis of VH107mR94 revealed a roughly 100-fold increase in affinity over VH107, while still maintaining lenalidomide-CRBN specificity, suggesting this optimized VH107mR94 could improve degradation of a CAR upon fusion (**Fig 3.5d-e**).

3.4 Discussion and Future Directions

Here, we demonstrate a strategy to use chemo-epitope specific sdAbs to degrade CARs in T cells. We began by generating these sdAbs against E3-ligases and small molecules complexes using a synthetic phage display library. From the top hits, several show remarkable specificities to the ternary complex. VH107, was the most specific clone by ELISA, but it was the only clone isolated against IMiD-CRBN. This is despite being able to bind all three IMiD moleucle, suggesting this clone should be present in the other two IMiD selections. Taken together these results suggest a bottlenecking of the selection and in the future deep sequencing of earlier round should be used to find other potential unique clones. Additionally, the small solvent accessible surface area of IMiDs in the ternary complex and the pan-IMiD recognition of VH107 suggest it is a conformational binder. Epitope binning with recombinant IKZF3, one of the known binders for IMiD-CRBN, would further validate this. Through these selections we also identified VH108 and several other binders to the small molecule binding pocket of cIAP, which were competitive with all SMAC mimetics tested. Interestingly, none of the CDR sequences share a similarity with the tetra-peptide motif for which the SMAC mimetics are based. We are uncertain how these sdAb binders came out of the selection, as small molecules were preincubated with E3 ligases before selections and kept in excess in buffer solutions. Additionally, ELISA signal ratios were not predictive of disrupter or dimerizer activity. In the future, a screening methodology like FACS and

sdAb yeast display could be used for more certainty during selections. This serendipitous discovery reinforces the pocket binding ability of sdAbs, to specifically enrich for disrupters sdAbs during the selection we envision elution of sdAb-phage from E3 ligase using a small molecule. While this would require high-affinity small molecules it would yield pocket-directed binders.

We subsequently tested VH107 and VH108 binding in cells using the SPPIER assay. While VH108 showed poor expression and no reversible binding, this tracks with the poor size-exclusion (SEC) profile and low thermostability (data not shown), suggesting further optimization of the molecule is needed for it to fold in the intracellular environment. VH107 showed binding in the SPPIER assay and was successfully fused to the CD19 CAR. Lenalidomide addition induced dose-dependent degradation of surface CAR, thus showing functional recruitment of endogenous CRBN and degradation machinery. Indeed, an exciting result, however, the dynamic range of CAR expression is relatively low. Coculture killing experiments are necessary to determine if lenalidomide-induced "ON" and "OFF" states can produce robust cytotoxicity in the former and no cytotoxicity in the latter. The dynamic range of CD19-IKZF3 was also low, c-terminal fusions to the CAR could generally destabilize the construct. Switching the extracellular anti-CD19 scfv to a different clone or sdAb could improve expression.

To improve the dynamic range of the CD19-VH107 CAR we attempted to improve the stability and affinity of VH107 through affinity maturation with yeast display. Excitingly, we identified an improved binder with better affinity and expression than the parental. Further replicates of kinetic characterization for this clone are needed to verify affinity constants. While the affinity increase of VH107mR94 is 100-fold greater than wt during this replicate, we observed much lower affinities of the parental VH107 than previous replicates (**Fig 3.5d** compared to **Fig 3.2c**). Despite this, we have confidence that VH107mR94 has improved affinity VH107 based on

our yeast flow cytometry comparisons (**Fig 3.5a**). VH107mR94 was the fourth most abundant clone in R8 but had the best affinity. The more abundant clones contained the same mutations in CDR H3 however had fewer or less mutations in H2 and the scaffold. The most abundant clone represented ~42% read abundance suggesting a jackpotting effect. Further maturation could be performed via AHEAD with VH107mR94 as a starting sequence. VH107mR94 will need to be tested as a CAR fusion to show that increase in affinity and expression will improve degradation of the CAR.

While numerous ways of regulating CAR T cell toxicities have been proposed, we believe our approach offers an additional benefit. Through our approach, multiple "ON" and "OFF" switches can be generated for the small molecule E3-ligase pairs of choice. These pairings are no longer limited to co-opted bacterial systems or repurposed E3-ligase substrates and can be derived from a human scaffold. Concerning limitations, the current CD19-VH107 needs to be fully validated in cytotoxicity assays and in mouse models to definitively prove degradation of CAR can abrogate toxic side effects. Overall, we hope this system is a step towards safer more accessible immunotherapies.

3.5 Materials and Methods

Cloning, Protein Expression, and Purification

The cereblon domain of unknown activity, binding cellular ligands and thalidomide (CULT), and the BIR3 domain of ciap were purchased as gblock fragments (IDT). These gene fragments contained N-terminal fusion of an AviTag and Tobacco Etch Virus (TEV) cut site. Gene fragments were then cloned into pMCSG7 vector [22]. Sequence verified plasmids were transformed into BL21 (λDE3) pBirA *E. coli* cells (Amid Biosciences). Single colonies were used to inoculate 1 L of 2xYT in carbenicillin (100μg/mL). Cultures were grown overnight at 37°C to an optical density of 0.8 at 600nm (OD_{600}) cooled to 18ºC, induced with a final concentration of 0.5mM IPTG, and grown for ~16hrs. Cells were pelleted and stored at -80ºC or purified directly.

Cell pellets were resuspended in 10mL of ice-cold lysis buffer (50 mM Tris, pH 8.0, 200 mM NaCl, 20 mM imidazole) supplemented with PMSF (100 μg/mL) and 10mL of B-PER (Thermo Fisher Scientific). Resuspended lysate was allowed to sit for 10 minutes on ice and cleared by centrifucation (20,000*g*, 20 min) at 4 °C. Cleared lysate were added to 200 μL of Ni-NTA Superflow resin (Qiagen) and rotated at 4 °C for 1 h. The resin was washed 3 times with lysis buffer and then transferred to a spin column. The purified protein was eluted with elution buffer (50 mM Tris, pH 8.0, 200 mM NaCl, 600 mM imidazole). Fractions were analyzed by SDS–PAGE, and exchanged into storage buffer (25 mM Tris, pH 8.0, 150 mM NaCl, 1 mM DTT).

VH domains were subcloned from the VH-phagemid into *E. coli* expression vector pBL347. Constructs were expressed in *E. coli* C43(DE3) Pro+ using optimized autoinduction medium and purified by protein A affinity chromatography as previously described [17].

Phage selection with VH-phage library

Phage selections were done according to previously established protocols [17]. Selections were performed using biotinylated CRBN or cIAP captured with streptavidin-coated magnetic beads (Promega). In each round, the phage pool was first cleared by incubation with beads loaded with 500 nM of the apoprotein. A second set of loaded beads were incubated with 2μM small molecule. The cleared phage were then incubated with this second set of beads containing the protein small molecule complex. After washing, the bound phage was eluted by the addition of 2 μg/mL of TEV protease. In total, four rounds of selection were performed with decreasing amounts of target protein. Steps for phage clearance were done in PBS buffer + 0.02% Tween-20 + 0.2% BSA (PBSTB). Steps for the positive selection were done in PBSTB + 2 μM small molecule. Individual phage clones from the third and fourth round of selections were analyzed by phage ELISA.

Phage ELISA

For each phage clone, four different conditions were tested—direct: protein with 2 μM small molecule; competition: protein with 2 μM small molecule with an equal concentration of protein in solution; negative selection: apoprotein; control: BSA. Flat-bottomed clear plates (Nunc MaxiSorp 384-well, Thermo Fisher Scientific) were coated with 0.5 μg ml^{-1} of NeutrAvidin in PBS overnight at 4 °C and subsequently blocked with PBS + 0.02% Tween-20 + 2% BSA for 1 h at room temperature. Plates were washed three times with PBS containing 0.05% Tween-20 (PBST) and were washed similarly between each of the steps, then 20 nM biotinylated protein was captured on the NeutrAvidin-coated wells for 30 min, then blocked with PBSTB + 10 μM biotin for 30 min. Phage supernatant diluted 1:5 in PBSTB was added for 20 min. For the competition

samples, the phage supernatant was diluted into PBSTB with 20 nM protein and small molecule. Bound phage were detected by incubation with anti-M13 antibody conjugated to horseradish-peroxidase (HRP) (Sino Biological; 1:5,000) for 30 min, followed by the addition of 3, 3', 5, 5' tetramethyl benzidine (TMB) substrate (VWR International). The reaction was quenched with the addition of 1 M phosphoric acid, then absorbance at 450 nm was measured using a Tecan M200 Pro spectrophotometer.

Biolayer interferometry

Bio-layer interferometry data (BLI) were measured using an Octet RED384 (ForteBio) instrument. Protein with small molecule or apoprotein were immobilized on a streptavidin and loaded until a 0.4 nm signal was achieved. After blocking with 10 μM biotin, purified binders in solution was used as the analyte. PBSTB was used for all buffers for BLI at pH 7.4. Small molecule was supplemented to a final concentration of 2 μM when appropriate. Data were analyzed using the ForteBio Octet analysis software and kinetic parameters were determined using a 1:1 monovalent binding model. Dose responses for Lenalidomide and Birinapant were performed using saturating concentrations of VH (VH107 at 900nM and VH108 at 450nM) and normalized to the maximum binding signals at 60 seconds during the association phase.

SPPIER assay

SdAbs and E3 ligases were incorporated into HOTag fusions as previously descried [19]. In brief, CRBN or cIAP1were first cloned into pcDNA3 containing EGFP. HOTag3 was then cloned into the pcDNA3 E3 ligase-EGFP construct, resulting in pcDNA3 E3-EGFP-HOTag3. Similar procedures were carried out to produce sdAb fusions. HEK293*T*/17 cells were passaged and

transiently transfected with the plasmids using lipofectamine. Cells were imaged 24 h after transient transfection. Time-lapse imaging was performed with the aid of an environmental control unit incubation chamber (InVivo Scientific). Fluorescence images were acquired with an exposure time of 50 ms for EGFP. Chemical reagents were carefully added to the cells in the incubation chamber when the time-lapse imaging started. Image acquisition was controlled by the NIS-Elements Ar Microscope Imaging Software (Nikon). Images were processed using NIS-Elements and ImageJ (NIH).

CAR Constructs All CAR constructs were of identical composition aside from the fusion of the intracellular degron. Overall, the fixed components are identical to the clinically approved CD-19 directed CAR construct tisangenlecleucel. Specifically, it is made up of an extracellular anti-CD19 scFv, a CD8 hinge and transmembrane domain, 4-1BB co-stimulatory domain, a CD3ζ signaling domain, and the degron. A T2A self-cleaving sequence and enhanced green fluorescent protein (eGFP) was inserted C-terminally to the degron, or CD3ζ domain in the degron-free construct, to identify CAR-positive cells. Similarly, a N-termial c-myc-tag was inserted on the anti-CD19 scFv for CAR-positive cell detection.

Transduction of Culture of Human T Cells

Primary human T cells were isolated from leukoreduction chamber residuals following Trima Apheresis (Blood Centers of the Pacific) using established protocols [23]. In brief, peripheral blood mononuclear cells (PBMCs) were isolated using Ficoll separation in SepMate tubes (STEMCELL Technologies) in accordance with the manufacturer's protocol. $CD8^+$ T cells were isolated from PBMCs using either the EasySep Human $CD8^+$ T cell Isolation Kit or the RosetteSep Human $CD8^+$ T Cell Enrichment Cocktail (STEMCELL), following the manufacturer's

protocol. T cells were cultured in ImmunoCult-XF T Cell Expansion media. (STEMCELL) and were passaged after 2-3 days. T cells were stimulated with CD3/CD28 Dynabeads (Thermo Fisher Scientific) according to the manufacturer's instructions for 5 days and supplemented with 30 U/mL recombinant interleukin 2 (ProSpec-Tany Technogene). Transduction with CAR lentivirus was performed 1 day after the start of bead stimulation. After the removal of activation beads, cells were optionally enriched by magnetic activated cell sorting and then expanded for another 4 days. CAR positivity was assessed by flow cytometry (GFP^+) and staining with anti-c-myc-647 (Cell Signaling). During CAR degradation experiments, media was supplemented with Lenalidomide or DMSO and incubated for 24hrs before CAR positivity was evaluated.

Affinity Maturation via Autonomous Hypermutation Yeast Surface Display

VH107 was integrated into the AHEAD yeast strain (yAW301) as previously described [21]. In brief, VH107 was subcloned into the pAW240 integration plasmid, linearized via restriction digest, and transformed into chemically competent yAW301 using Frozen-EZ Yeast Transformation II kit (Zymo). After 3 days of growth on selective SC-HLUWMC dropout media (US Biologics) single colonies were grown to saturation SC-HLUWMC supplemented with 2% glucose. Successful VH107 incorporation was validated by sanger sequencing and further diversified by 2-3 1000x fold dilutions into SC-HLUW with 2% glucose. For FACS, VH107 expression was first induced in yeast by switching to SC-HLUW with 2% galactose for 48hrs. Yeast were labeled with biotinylated CRBN for 1-4hr, with round 1 starting at 50nM and every subsequent round at half the previous concentration. Yeast cell counts, incubation volumes, and times were further adjusted to prevent ligand depletion and allow binding to reach equilibrium. After washing, yeast were stained with 1000x dilution of anti-HA-AF488 (Thermo cat. no. 26183),

washed again and subjected to FACS. During each round ~2e7 cells were stained and used to sort out 200-2,000 cells. To prevent mutants with HA-tag mutations from passing a strict floor on HA signal was set to include only the top ~15% of clones on HA signal axis. Cells were sorted into 3mL of SC-HLUW with 2% glucose and allowed to grow to saturation (~3 days), whereby they were subjected to the next round.

3.6 Main Figures

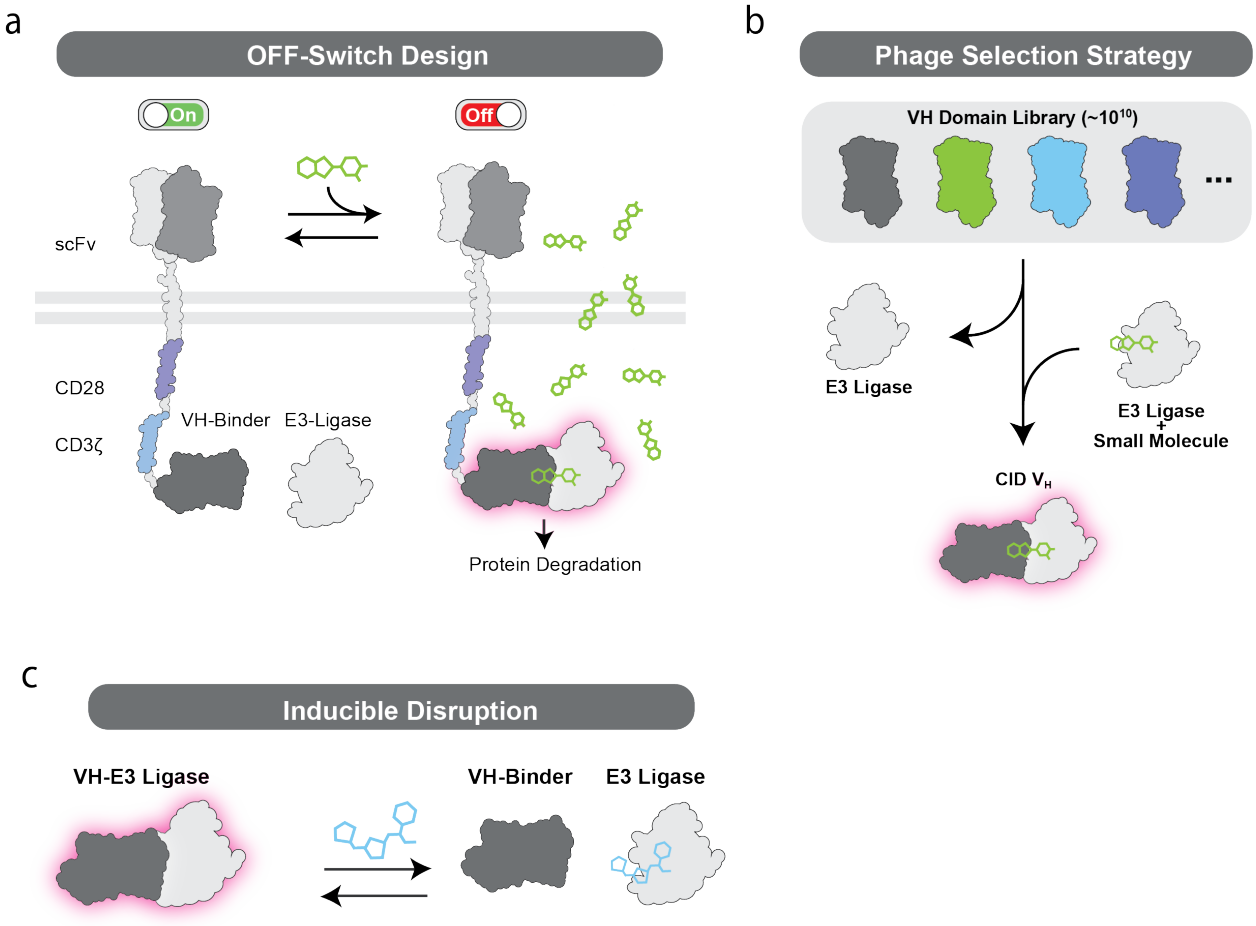


Figure 3.1: Design strategy and application of sdAb based switchs
(**A**) Schematic of degron based OFF-switch for CAR-T cells (**B**) Phage selection strategy for creating chemically inducible dimerizers (CIDs) between sdAbs and E3 ligases (**C**) Schematic for chemically inducible disrupters that were discovered. The small molecule (blue) is capable of disrupting interaction between sdAbs and E3 ligase.

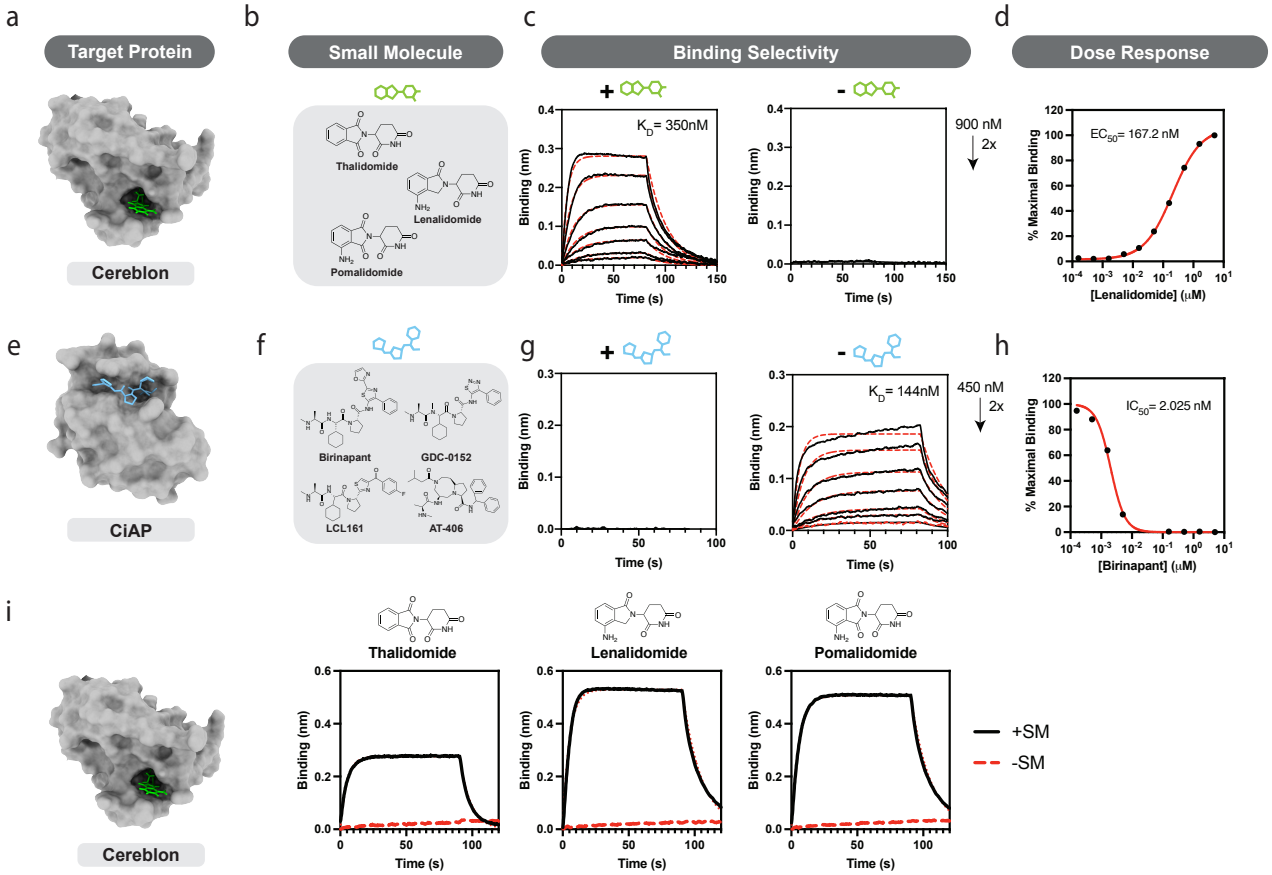


Figure 3.2: Characterization of top sdAb binders

(**A**) CULT domain of CRBN shown with pomalidomide bound, PDB: 6H0F (**B**) Chemical structure of IMiDs used in phage selection, represented as green cartoon (**C**) Biolayer Interferometry (BLI) of VH107 +lenalidomide (left) and -lenalidomide (right) (**D**) BLI dose titration of lenalidomide using saturating VH107 (900nM) (**E**) BIR domain of cIAP1 shown with GDC-0152 bound, PDB: 3UW4 (**F**) Chemical structure of SMAC mimetics, represented as blue cartoon (**G**) BLI of VH108 +birinapant (left) and -biriniapant (right) (**H**) BLI dose titration of birinapant using saturating VH107 (450nM) (**I**) BLI of VH107 against all IMiDs shows productive binding only in the presence of small molecule.

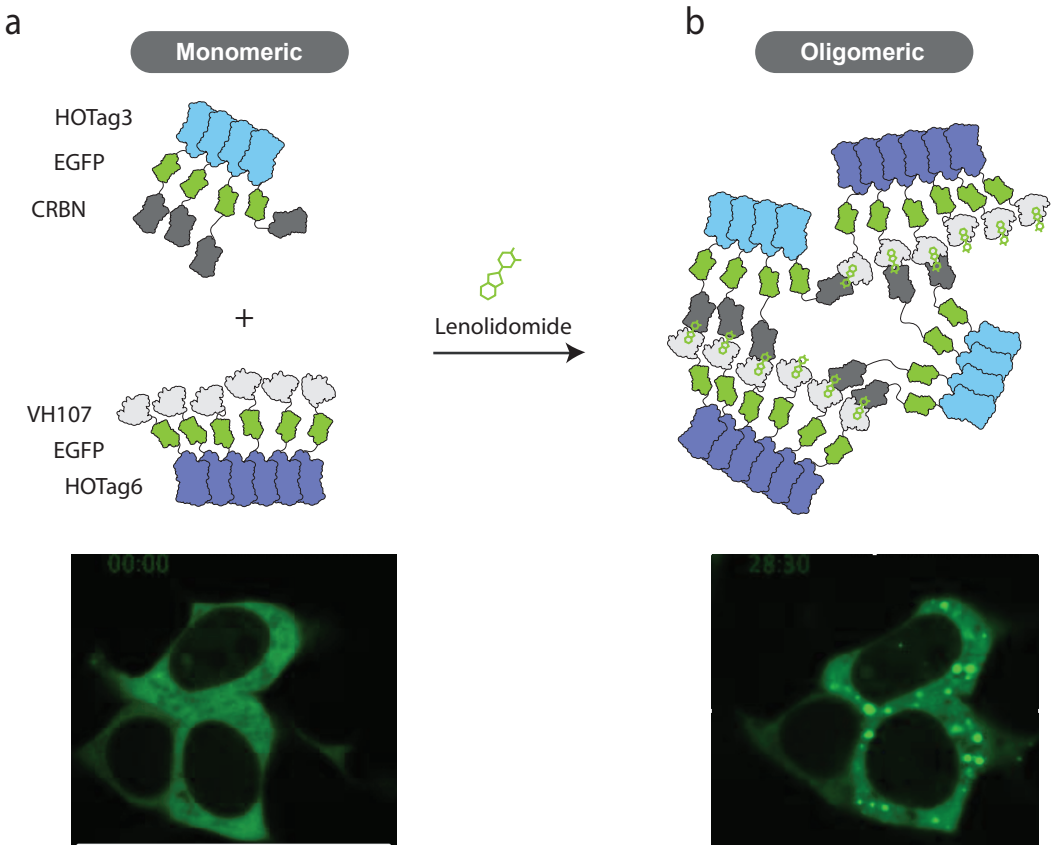


Figure 3.3: SPPIER assay reveals ternary complex formation
Schematic detailing monomeric HOTag3 and HOTag6 fusions to VH107 and CRBN in monomeric and oligomeric form (**A**) Microscopy image shows cells before lenalidomide addition with low EGFP fluorescence. (**B**) Microscopy image shows cells at final time point after lenalidomide addition with phase separated puncta of high EGFP fluorescence.

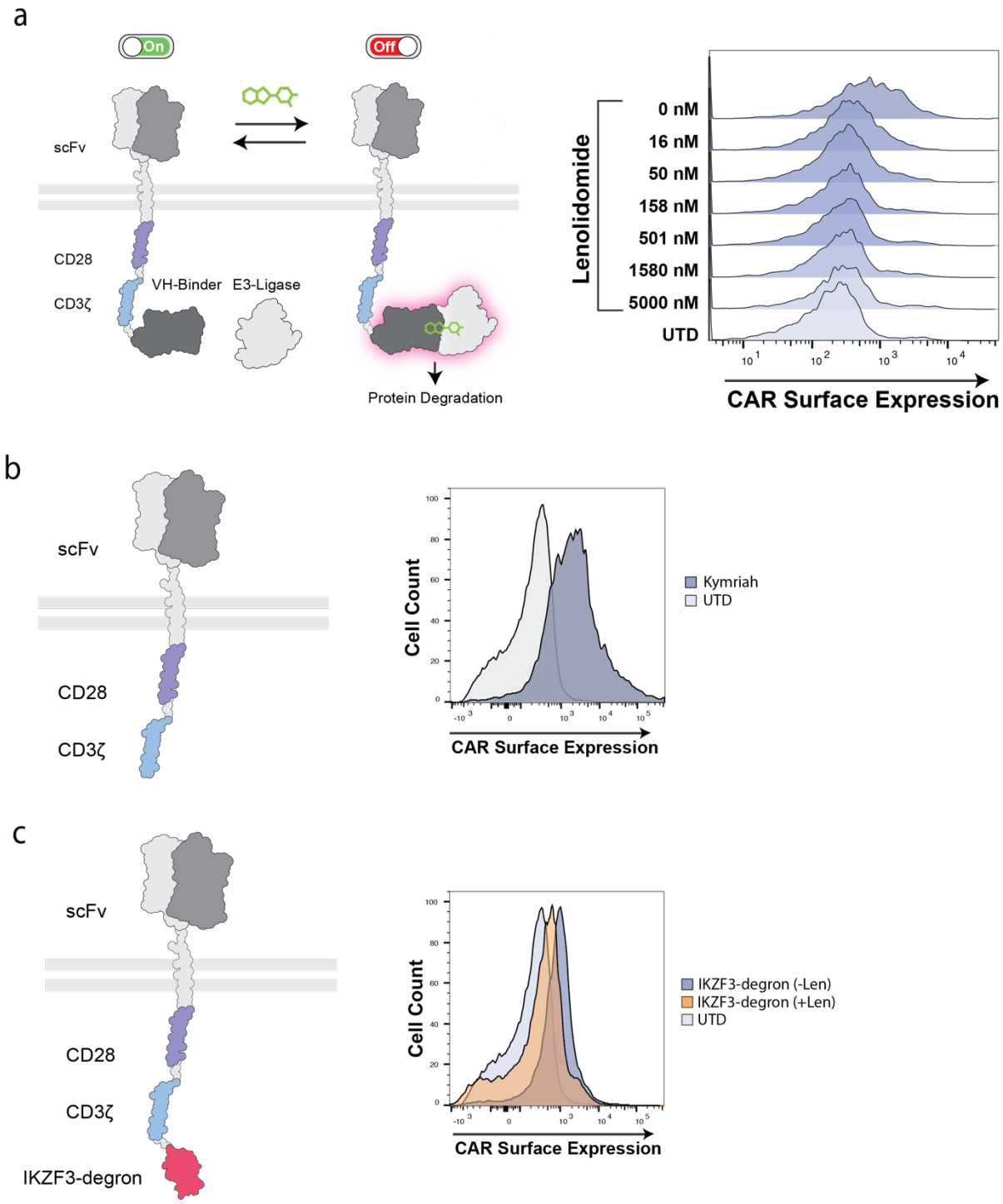


Figure 3.4: VH107 OFF-switch enables degradation of CAR
Cartoon representation of each CAR architecture (left). Flow cytometric analysis of surface CAR expression for (**A**) CD19-VH107 CAR, (**B**) CD19 CAR, and (**C**) CD19-IKZF3 CAR (right). Analysis for CD19 CAR - lenalidomide is shown here, however +/- lenalidomide conditions were tested and surface level CAR expression was comparable.

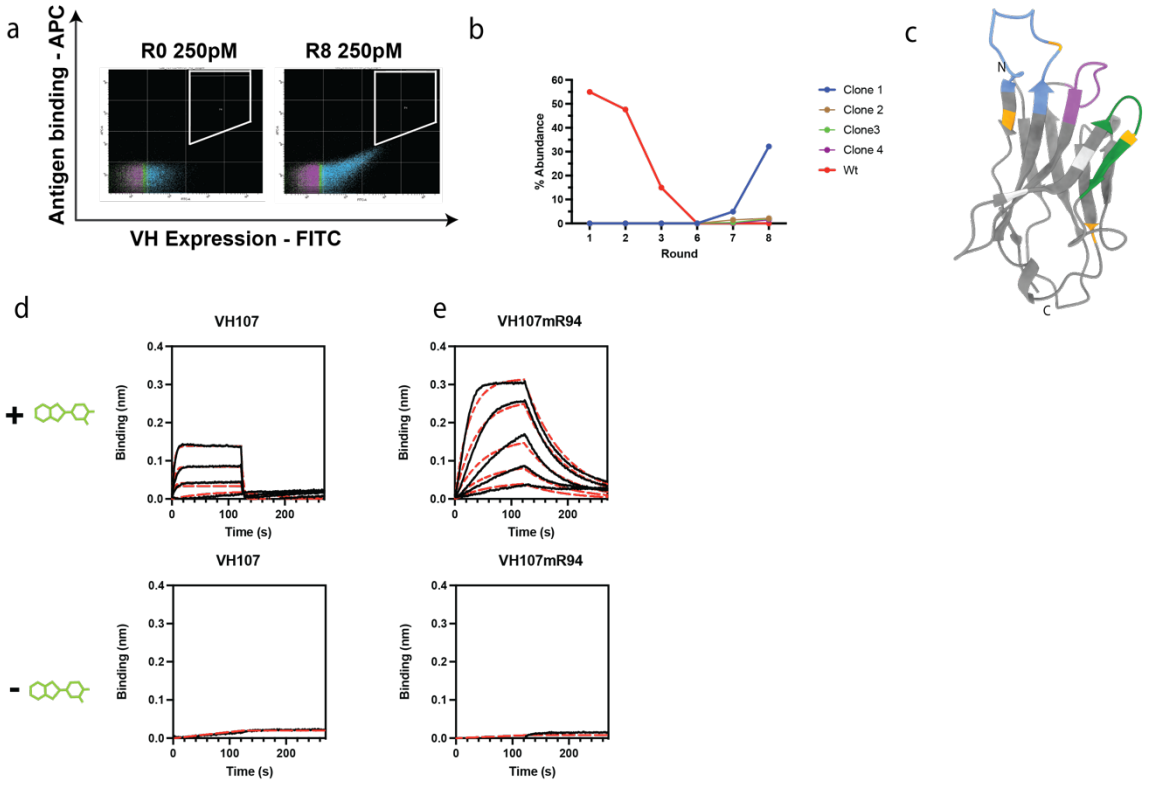


Figure 3.5: Affinity maturation of VH107

(**A**) FACs plots show for the VH107 parental clone (left) and for R8 output library (right) both stained at 250pM CRBN. (**B**) Next-generation sequencing (NGS) read abundance of VH107 parental and R8 top clones tracked over R1-R3 and R6-R8. (**C**) Structural positions of mutations in 4 locations (orange) of VH107mR94. Positions in light grey denote scaffold mutations that were present in most abundant clone. **(D)** Multipoint BLI of VH107 (starting at 450nM with subsequent 2x dilutions) with and without lenalidomide (5μM). **(E)** BLI of VH107mR94 with and without lenalidomide.

3.7 Extended and Supplemental Figures

Table 3.S1: Differential Scanning Fluorimetry (DSF) of VH107
The melting temperature (Tm) of each VH binder was determined by differential scanning fluorimetry.

VH ID	Tm C°
VH107	58.8
VH107 + 2.5% BME	47.9

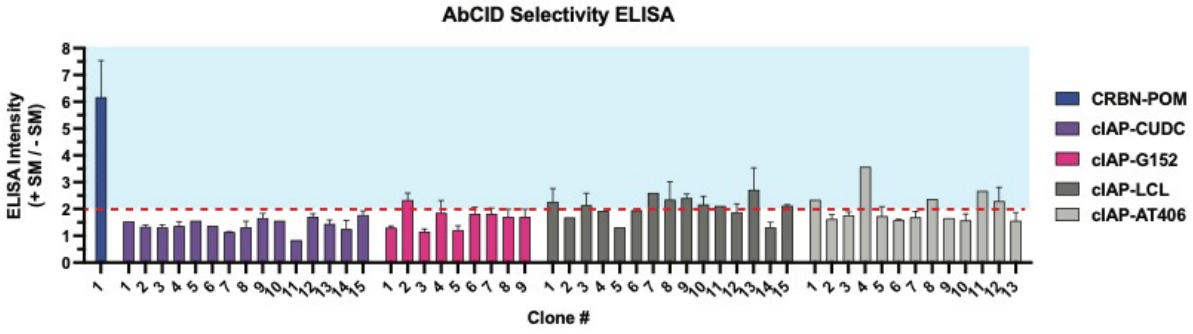


Figure 3.S1: Characterization of unique clones from sdAb library
Plot of ELISA signal as a ratio between E3-ligase +SM and E3-ligase -SM. Intensity of signal is average of all unique clones from sequencing ~400 total clones. A cutoff ratio of two was used to triage hits specific to E3-ligase +SM. Passing clones reach the blue threshold.

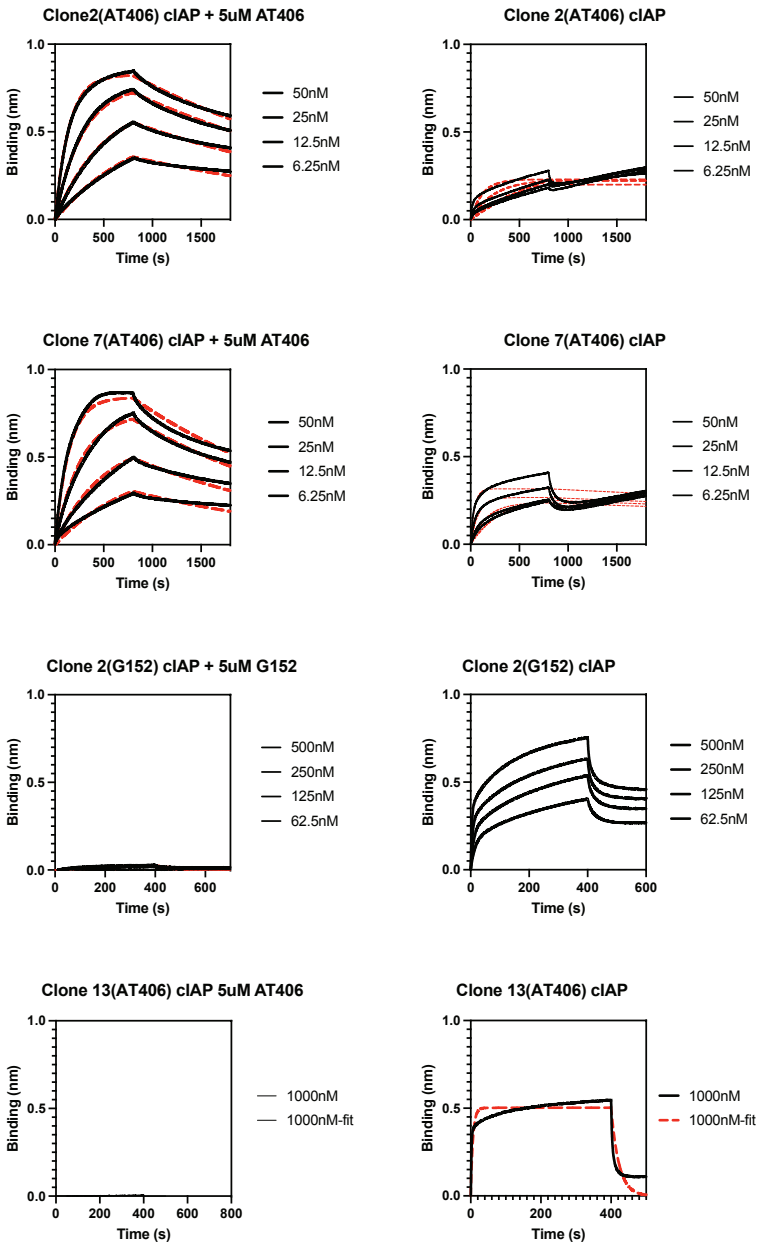


Figure 3.S2: Biolayer Interferometry (BLI) of select sdAbs clones
Multipoint and singlepoint BLI experiments for top clones binding to cIAP1 with two SMAC mimetics. Two binders, to GDC-0152 and AT406, bind only in absence of small molecule and were classified as chemically inducible disrupters. Overall, affinities were not determined.

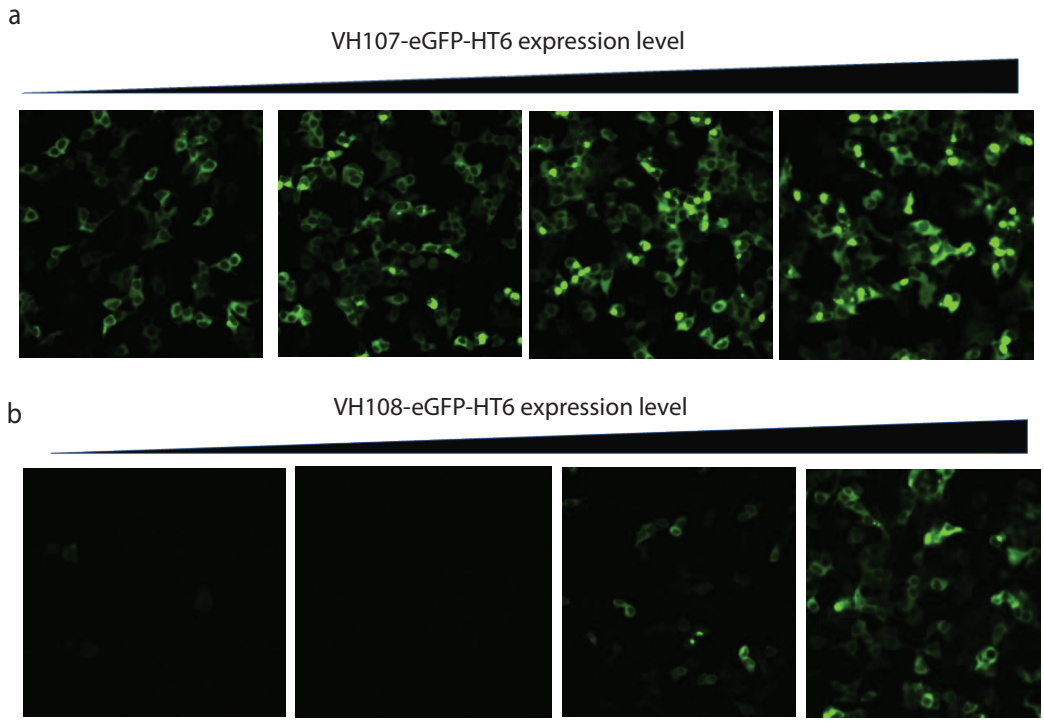


Figure 3.S3: Optimization of SPPIER assay
Representative micrographs of SPPIER assay at increasing levels of (**A**) VH107-eGFP-HT6 and (**B**) VH108-eGFP-HT6. VH108 shows poor expression as fusion.

3.8 References

1. Cohen, Adam D., et al. "B cell maturation antigen–specific CAR T cells are clinically active in multiple myeloma." *The Journal of clinical investigation* 129.6 (2019): 2210-2221.
2. Davila, Marco L., et al. "Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia." *Science translational medicine* 6.224 (2014): 224ra25-224ra25.
3. Schuster, Stephen J., et al. "Chimeric antigen receptor T cells in refractory B-cell lymphomas." *New England Journal of Medicine* 377.26 (2017): 2545-2554.
4. Bonifant, Challice L., et al. "Toxicity and management in CAR T-cell therapy." *Molecular Therapy-Oncolytics* 3 (2016): 16011.
5. Straathof, Karin C., et al. "An inducible caspase 9 safety switch for T-cell therapy." *Blood* 105.11 (2005): 4247-4254.
6. Vogler, Isabel, et al. "An improved bicistronic CD20/tCD34 vector for efficient purification and in vivo depletion of gene-modified T cells for adoptive immunotherapy." *Molecular Therapy* 18.7 (2010): 1330-1338.
7. Wu, Chia-Yung, et al. "Remote control of therapeutic T cells through a small molecule–gated chimeric receptor." *Science* 350.6258 (2015): aab4077.
8. Rodgers, David T., et al. "Switch-mediated activation and retargeting of CAR-T cells for B-cell malignancies." *Proceedings of the National Academy of Sciences* 113.4 (2016): E459-E468.
9. Carbonneau, Seth, et al. "An IMiD-inducible degron provides reversible regulation for chimeric antigen receptor expression and activity." *Cell chemical biology* 28.6 (2021): 802-812.

10. Jan, Max, et al. "Reversible ON-and OFF-switch chimeric antigen receptors controlled by lenalidomide." *Science translational medicine* 13.575 (2021): eabb6295.

11. Krönke, Jan, et al. "Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells." *Science* 343.6168 (2014): 301-305.

12. Lu, Gang, et al. "The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins." *Science* 343.6168 (2014): 305-309.

13. Sperling, Adam S., et al. "Patterns of substrate affinity, competition, and degradation kinetics underlie biological activity of thalidomide analogs." *Blood, The Journal of the American Society of Hematology* 134.2 (2019): 160-170

14. Sievers, Quinlan L., et al. "Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN." *Science* 362.6414 (2018): eaat0572.

15. Hill, Zachary B., et al. "Human antibody-based chemically induced dimerizers for cell therapeutic applications." *Nature chemical biology* 14.2 (2018): 112-117.

16. Amaravadi, Ravi K., et al. "A Phase I Study of the SMAC-Mimetic Birinapant in Adults with Refractory Solid Tumors or LymphomaPhase I Trial of Birinapant." *Molecular cancer therapeutics* 14.11 (2015): 2569-2575.

17. Bracken, Colton J., et al. "Bi-paratopic and multivalent VH domains block ACE2 binding and neutralize SARS-CoV-2." *Nature chemical biology* 17.1 (2021): 113-121.

18. Matyskiela, Mary E., et al. "A cereblon modulator (CC-220) with improved degradation of Ikaros and Aiolos." *Journal of medicinal chemistry* 61.2 (2018): 535-542.

19. Chung, Chan-I., Qiang Zhang, and Xiaokun Shu. "Dynamic imaging of small molecule induced protein–protein interactions in living cells with a fluorophore phase transition based approach." *Analytical chemistry* 90.24 (2018): 14287-14293.

20. Guo, Wen-Hao, et al. "Enhancing intracellular accumulation and target engagement of PROTACs with reversible covalent chemistry." *Nature communications* 11.1 (2020): 1-16.

21. Wellner, Alon, et al. "Rapid generation of potent antibodies by autonomous hypermutation in yeast." *Nature chemical biology* 17.10 (2021): 1057-1064.

22. Seiler, Catherine Y., et al. "DNASU plasmid and PSI: Biology-Materials repositories: resources to accelerate biological research." *Nucleic acids research* 42.D1 (2014): D1253-D1260.

23. Byrnes, James R., et al. "Hypoxia Is a Dominant Remodeler of the Effector T Cell Surface Proteome Relative to Activation and Regulatory T Cell Suppression." *Molecular & Cellular Proteomics* 21.4 (2022).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Colton Bracken
1F694405AF09416...
Author Signature

8/13/2022
Date