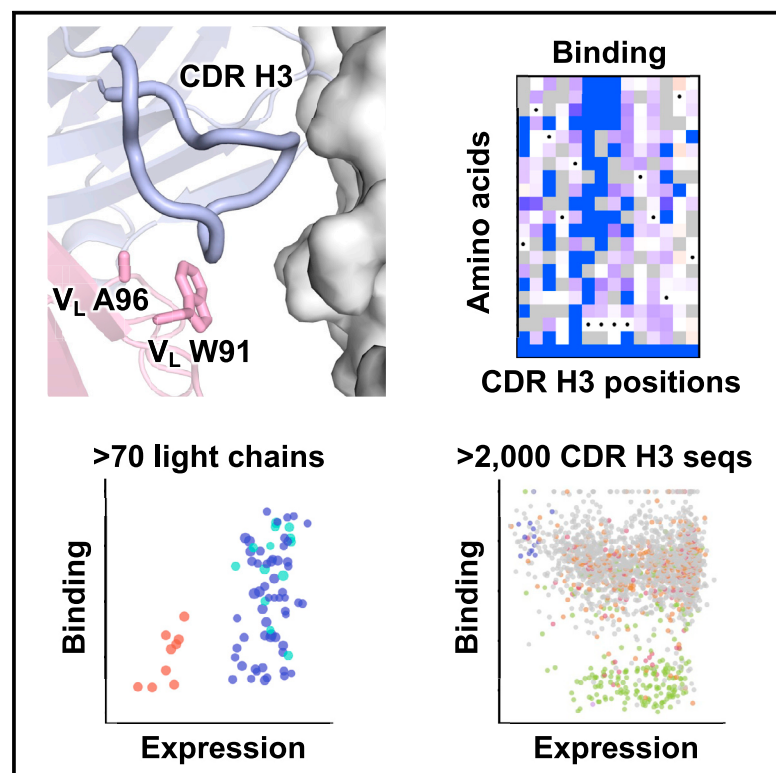


# Stringent and complex sequence constraints of an IGHV1-69 broadly neutralizing antibody to influenza HA stem

## Graphical abstract



## Authors

Qi Wen Teo, Yiquan Wang, Huibin Lv, Timothy J.C. Tan, Ruipeng Lei, Kevin J. Mao, Nicholas C. Wu

## Correspondence

nicwu@illinois.edu

## In brief

Through high-throughput experiments, Teo et al. systematically analyze the sequence constraints of CR9114, which is the broadest influenza antibody known to date. This work provides molecular insights into broadly neutralizing antibody responses to influenza virus, which have important implications for universal influenza vaccine development.

## Highlights

- Compatibilities of CR9114 with diverse light chains and CDR H3s are probed
- Light chain of CR9114, despite having no contact with the epitope, is important for binding
- Most CDR H3s from other IGHV1-69 HA stem antibodies are incompatible with CR9114



## Article

# Stringent and complex sequence constraints of an IGHV1-69 broadly neutralizing antibody to influenza HA stem

Qi Wen Teo,<sup>1,2,5</sup> Yiquan Wang,<sup>1,5</sup> Huibin Lv,<sup>1,2,5</sup> Timothy J.C. Tan,<sup>3</sup> Ruipeng Lei,<sup>1</sup> Kevin J. Mao,<sup>1</sup> and Nicholas C. Wu<sup>1,2,3,4,6,\*</sup>

<sup>1</sup>Department of Biochemistry, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

<sup>2</sup>Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

<sup>3</sup>Center for Biophysics and Quantitative Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

<sup>4</sup>Carle Illinois College of Medicine, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

<sup>5</sup>These authors contributed equally

<sup>6</sup>Lead contact

\*Correspondence: [nicwu@illinois.edu](mailto:nicwu@illinois.edu)

<https://doi.org/10.1016/j.celrep.2023.113410>

## SUMMARY

IGHV1-69 is frequently utilized by broadly neutralizing influenza antibodies to the hemagglutinin (HA) stem. These IGHV1-69 HA stem antibodies have diverse complementarity-determining region (CDR) H3 sequences. Besides, their light chains have minimal to no contact with the epitope. Consequently, sequence determinants that confer IGHV1-69 antibodies with HA stem specificity remain largely elusive. Using high-throughput experiments, this study reveals the importance of light-chain sequence for the IGHV1-69 HA stem antibody CR9114, which is the broadest influenza antibody known to date. Moreover, we demonstrate that the CDR H3 sequences from many other IGHV1-69 antibodies, including those to the HA stem, are incompatible with CR9114. Along with mutagenesis and structural analysis, our results indicate that light-chain and CDR H3 sequences coordinately determine the HA stem specificity of IGHV1-69 antibodies. Overall, this work provides molecular insights into broadly neutralizing antibody responses to influenza virus, which have important implications for universal influenza vaccine development.

## INTRODUCTION

Influenza viruses pose a constant threat to public health, resulting in substantial morbidity and mortality with approximately 500,000 deaths worldwide each year.<sup>1</sup> Although vaccination remains the foremost measure for preventing and controlling influenza virus infection, the effectiveness of the annual seasonal influenza vaccine has varied widely, ranging from 19% to 60% over the past decade.<sup>2</sup> This variability is largely due to the continuous antigenic drift of human influenza virus, especially at the immunodominant head domain of hemagglutinin (HA), which can in turn lead to vaccine mismatch in some influenza seasons.<sup>3,4</sup> In addition, the current seasonal influenza vaccine is designed to protect against influenza A H1N1 and H3N2 subtypes, as well as influenza B virus, but not avian influenza A subtypes with zoonotic potential, such as H5N1 and H7N9. Therefore, significant efforts have been made to develop a universal influenza vaccine that targets the highly conserved HA stem domain,<sup>5,6</sup> with an ultimate goal to elicit broadly neutralizing HA stem antibodies like CR9114, which can bind to both group 1 (e.g., H1 and H5) and group 2 (e.g., H3 and H7) HAs as well as influenza B HAs.<sup>7</sup>

Many HA stem antibodies, including CR9114, are encoded by IGHV1-69.<sup>8–10</sup> Structural analyses have shown that the paratopes of these IGHV1-69 antibodies are dominated by the heavy chain

with minimal to no contribution from the light chain.<sup>7,11–13</sup> Consistently, diverse light-chain germline genes are found among IGHV1-69 HA stem antibodies (e.g., IGKV1-44,<sup>7</sup> IGLV1-51,<sup>9,12</sup> IGLV10-54,<sup>11</sup> and IGKV3-20<sup>13</sup>). Similarly, the complementarity-determining region (CDR) H3 sequences, which are formed by VDJ recombination, are highly diverse among IGHV1-69 HA stem antibodies.<sup>14</sup> Although most IGHV1-69 HA stem antibodies encode a Tyr in the CDR H3, the position of this Tyr varies, and some do not even have a Tyr in the CDR H3.<sup>14</sup> Based on these observations, IGHV1-69 HA stem antibodies do not seem to have strong sequence preferences in the CDR H3 and light chain. At the same time, it is impossible that all IGHV1-69 antibodies can bind to the HA stem, given that many IGHV1-69 antibodies are specific to other pathogens.<sup>15</sup> Besides, some IGHV1-69 antibodies bind to the HA head instead of the HA stem.<sup>16,17</sup> Therefore, despite the first IGHV1-69 HA stem antibody being discovered 15 years ago,<sup>9</sup> it remains unclear what sequence features make an IGHV1-69 antibody bind to the HA stem.

In this study, we performed two high-throughput experiments to probe the sequence constraints in the light chain and CDR H3 of the IGHV1-69 HA stem antibody CR9114, which is the broadest influenza neutralizing antibody known to date.<sup>7</sup> Our first high-throughput experiment examined the compatibility of 78 light-chain sequences from diverse IGHV1-69 antibodies with



CR9114 heavy chain. Our findings indicated that despite having no contact with the HA stem epitope, the light chain of CR9114 contained sequence determinants for its binding activity. Specifically, we demonstrated that the amino acid sequences at variable light chain domain ( $V_L$ ) residues 91 and 96 hugely influenced the compatibility of light chain with CR9114 heavy chain for binding to the HA stem. Of note, the Kabat numbering scheme is used throughout. Our second high-throughput experiment measured the binding affinity of 2,162 diverse CDR H3 variants to the HA stem and showed that most CDR H3 variants, including many from other IGHV1-69 HA stem antibodies, were incompatible with CR9114. These results indicate that the sequence constraints in the CDR H3 and light chain of IGHV1-69 HA stem antibodies are stringent yet complex.

## RESULTS

### Light-chain sequence of CR9114 is important for HA stem binding

A previous structural study has shown that CR9114, which is encoded by IGHV1-69 and IGLV1-44, does not use light chain for binding (Figure 1A).<sup>7</sup> In addition, the HA-stem-binding activity of CR9114 is not affected by replacing its light chain with one from an antibody of different specificity.<sup>10</sup> Here, we aimed to systematically investigate if the sequence of CR9114 light chain is truly unimportant for HA stem binding. Briefly, we compiled a list of 120 antibodies from GenBank that are encoded by IGHV1-69 with different  $\lambda$  light chains (Table S1). The light-chain sequences from these 120 antibodies were synthesized and paired with the germline sequence of CR9114 heavy chain to create a light-chain variant library. We also included 10 light-chain sequences with premature stop codons as negative controls. As a result, our light-chain variant library contained 130 different light-chain sequences, all of which paired with the germline sequence of CR9114 heavy chain. The germline sequence of CR9114 heavy chain was used because we wanted to avoid any incompatibility between heavy-chain somatic hypermutations (SHMs) and light-chain variants. This heavy-chain germline sequence of CR9114 was reconstructed in a previous study,<sup>18</sup> which reverted the SHMs on V and J genes to germline sequences.

Subsequently, this light-chain variant library was displayed on the yeast cell surface and two-way sorted based on antibody expression level as well as binding activity to mini-HA, which is a trimeric HA stem construct without the head domain (Figure 1B).<sup>5</sup> Four sorted populations were collected, namely no expression, high expression, no binding, and high binding. The frequency of each light-chain variant in each sorted population was quantified by next-generation sequencing. Among the 130 light-chain variants in the library, 78 had an average occurrence frequency of >0.2% across different sorted populations and were subjected to downstream analyses (Table S1). These 78 light-chain variants include the one from CR9114 (i.e., wild type [WT]), 13 from other influenza antibodies, 55 from non-influenza antibodies, and 9 negative controls with premature stop codons. For each light-chain variant, an expression score and a binding score were computed based on its frequency in different sorted populations (see STAR Methods). Both the expression

and binding scores were normalized such that the scores for WT equaled 1 and the mean scores for negative controls (i.e., variants with stop codons) equaled 0. Pearson correlation coefficients of 0.75 and 0.74 were obtained between two replicates of the binding and expression sorts, respectively (Figure S1), confirming the reproducibility of our results.

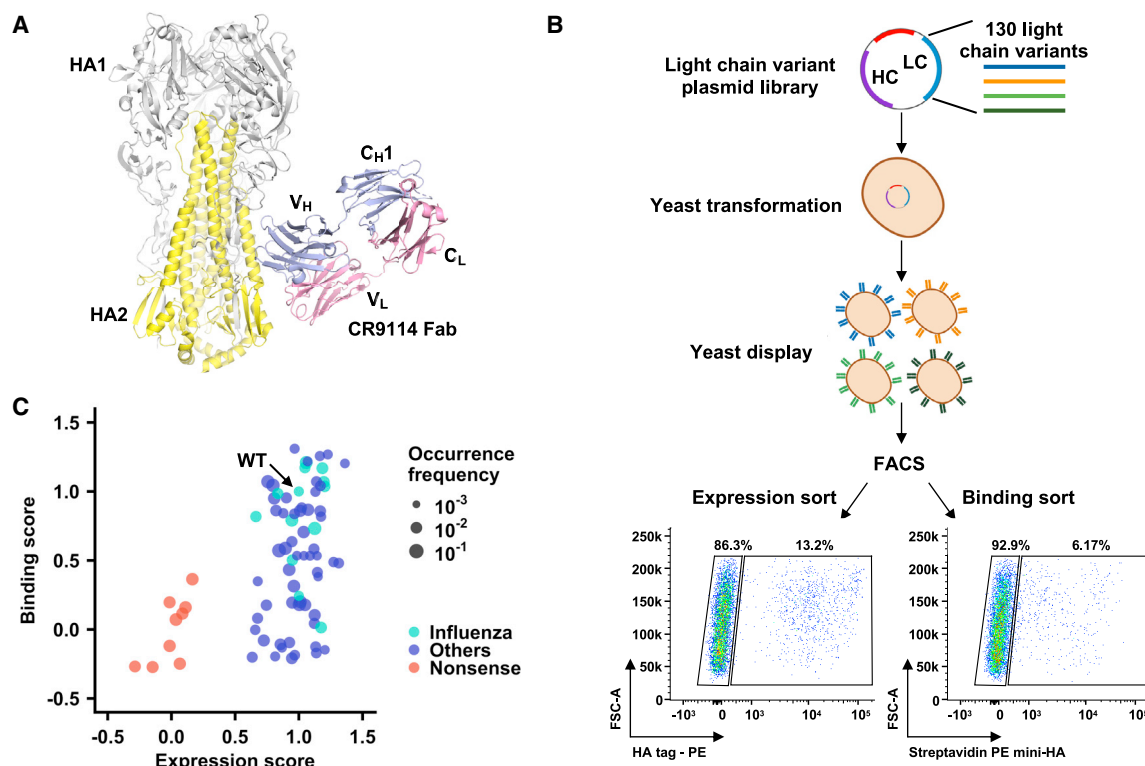
Except for the negative controls, most light-chain variants had an expression score of around 1 (Figure 1C), indicating that the light-chain sequence of the CR9114 germline had minimal influence on expression. However, many light-chain variants had a low binding score (Figure 1C). In addition, the binding scores of light-chain variants from influenza antibodies (mean = 0.84) were significantly higher than those from non-influenza antibodies (mean = 0.49,  $p = 0.005$ , two-tailed Student's  $t$  test), despite having similar expression scores (mean = 1.03 and 0.99, respectively,  $p = 0.35$ , two-tailed Student's  $t$  test). These results imply that the light-chain sequence of the CR9114 germline is an important determinant for its HA-stem-binding activity.

### Importance of CR9114 light-chain residues 91 and 96

Next, we aimed to understand the molecular mechanism of how the light chain modulated the HA-stem-binding activity of the CR9114 germline. Since CR9114 light chain has no contact with the HA stem epitope (Figure 1A),<sup>7</sup> its sequence determinants likely locate at the heavy-light chain interface. Structural analysis of the previously determined crystal structure of CR9114 in complex with HA<sup>7</sup> suggested that  $V_L$  residues 91 and 96 in CDR L3 are critical for stabilizing the conformation of CDR H3, which in turn is important for binding (Figure 2A). CR9114 has an aromatic residue Trp at  $V_L$  residue 91, which forms an extensive  $\pi$ - $\pi$  stacking network with four aromatic residues in CDR H3, namely variable heavy chain domain ( $V_H$ ) Y98,  $V_H$  Y99,  $V_H$  Y100, and  $V_H$  Y100a, as well as  $V_H$  W47 in the heavy-chain framework region 2. In contrast, CR9114 has a small amino acid Ala at  $V_L$  residue 96, which points toward the heavy chain with limited space in between. These observations suggested that the compatibility of different light-chain variants and the CR9114 germline heavy chain depended on the amino acid identities at  $V_L$  residues 91 and 96. Specifically, we hypothesized that amino acids F/Y/W, which are both aromatic and bulky, were required at  $V_L$  residue 91 but forbidden at  $V_L$  residue 96.

We then compared the amino acid sequences at  $V_L$  residues 91 and 96 between high-binding and low-binding light-chain variants using an arbitrary binding score cutoff of 0.8 (Figure 2B). At  $V_L$  residue 91, all light-chain variants with a binding score >0.8 contained an aromatic amino acid (i.e., W/Y/F), whereas non-aromatic amino acids could be observed among variants with a binding score <0.8. Conversely, aromatic amino acids were enriched at  $V_L$  residue 96 among variants with a binding score <0.8. Specifically, Trp and Tyr exhibited 5- and 4-fold enrichments, respectively, whereas Phe was absent among variants with a binding score >0.8 but present at 5% among variants with a binding score >0.8 (Figure 2B; Table S2). These observations are consistent with our hypothesis above.

To further experimentally validate our findings, we introduced different mutations at residues 91 and 96 of CR9114 light chain, recombinantly expressed them by pairing them with the CR9114 germline heavy chain, and then tested their binding affinity to



**Figure 1. Yeast display of light-chain variant library with CR9114 germline heavy chain**

(A) Interaction between HA and CR9114 (PDB: 4FQI).<sup>7</sup> Gray: HA1; yellow: HA2; blue: heavy chain; pink: light chain. C<sub>H</sub>1 indicates constant domain 1 of heavy chain, and C<sub>L</sub> indicates constant domain 1 of light chain. V<sub>H</sub> and V<sub>L</sub> indicate variable domains of antibody heavy and light chains, respectively. (B) Schematic illustration of measuring the expression level and HA-stem-binding activity of many light-chain variants in parallel using yeast surface display. Briefly, an antibody light-chain (LC) variant library was paired with CR9114 germline heavy chain (HC), displayed on the yeast cell surface, and subjected to fluorescence-activated cell sorting (FACS) based on surface expression level and binding to mini-HA. The sorted populations were analyzed by next-generation sequencing. Two independent replicates were performed.

(C) Relationship between binding and expression scores. One datapoint represents one light-chain variant. The datapoint corresponding to CR9114 WT light chain is labeled as "WT." Cyan: light-chain variants from known IGHV1-69 antibodies to influenza virus. Blue: light-chain variants from IGHV1-69 antibodies to non-influenza antigens. Red: negative control variants with stop codons.

See also [Figure S1](#) and [Table S1](#).

mini-HA. Hereafter, CR9114 with germline heavy chain is abbreviated CR9114 gHC. The binding activity of CR9114 gHC to mini-HA was abolished by substituting V<sub>L</sub> W91 with non-aromatic amino acids T/R/A but not aromatic amino acids Y/F ([Table 1](#); [Figure S2A](#)). Likewise, the binding activity of CR9114 gHC to mini-HA was abolished by substituting V<sub>L</sub> A96 with aromatic amino acids W/F but not non-aromatic amino acids V/S/R ([Table 1](#); [Figure S2A](#)). We also assessed the binding activity of these CR9114 light-chain mutants to the HA protein from H1N1 A/Solomon Islands/3/2006 (SI06) and observed a similar trend ([Table 1](#); [Figure S2B](#)). Therefore, our binding experiment confirms that V<sub>L</sub> residues 91 and 96 in the CDR L3 are important for CR9114 to interact with the HA stem, despite not being part of the paratope.

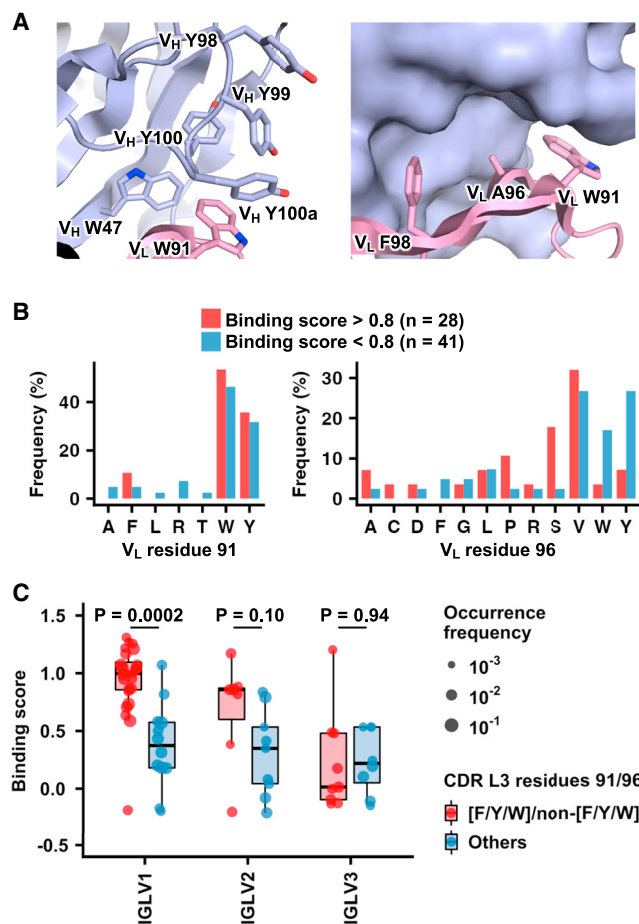
#### Additional sequence constraints in CR9114 light chain

Although our results indicated that aromatic amino acids were required at V<sub>L</sub> residue 91 but forbidden at V<sub>L</sub> residue 96 of CR9114, exceptions existed in our light-chain variant library

([Table S1](#)). Specifically, we noticed that some light-chain variants with a low binding score had aromatic (i.e., F/Y/W) and non-aromatic amino acids (i.e., non-F/Y/W) at V<sub>L</sub> residues 91 and 96, respectively (i.e., V<sub>L</sub> 91<sub>F/Y/W</sub>/96<sub>non-F/Y/W</sub>). This observation suggests that there are additional sequence constraints in the light chain of CR9114.

Members of our light-chain variant library were from three IGLV families, namely IGLV1, IGLV2, and IGLV3. For light-chain variants in the IGLV1 family, those with V<sub>L</sub> 91<sub>F/Y/W</sub>/96<sub>non-F/Y/W</sub> had significantly higher binding scores than those without ( $p = 2e-4$ , two-tailed Student's *t* test; [Figure 2C](#)). In contrast, while the light-chain variants in the IGLV2 family with V<sub>L</sub> 91<sub>F/Y/W</sub>/96<sub>non-F/Y/W</sub> also had higher binding scores than those without, such a difference was not as significant ( $p = 0.10$ , two-tailed Student's *t* test; [Figure 2C](#)). Furthermore, the binding scores of light-chain variants in the IGLV3 family with and without the V<sub>L</sub> 91<sub>F/Y/W</sub>/96<sub>non-F/Y/W</sub> motif had no significant difference ( $p = 0.94$ , two-tailed Student's *t* test) and were generally low ([Figure 2C](#)). Of note, in all three IGLV families, the expression scores





**Figure 2. Light-chain sequence determinants for HA-stem-binding activity of CR9114**

(A) Left: an extensive  $\pi$ - $\pi$  stacking network at the heavy-light chain interface of CR9114 that involves V<sub>H</sub> W47, V<sub>H</sub> Y98, V<sub>H</sub> Y99, V<sub>H</sub> Y100, V<sub>H</sub> Y100a, and V<sub>L</sub> W91. Right: side chains of V<sub>L</sub> W91, A96, and F98 at the heavy-light chain interface of CR9114, with heavy chain in surface representation. PDB: 4FQI is used. Light blue: heavy chain; pink: light chain. V<sub>H</sub> and V<sub>L</sub> indicate variable domains of antibody heavy and light chains, respectively.

(B) Frequency of amino acid usage at V<sub>L</sub> residues 91 and 96 of light-chain variants with binding scores > 0.8 (red) and < 0.8 (blue) is shown.

(C) Binding scores of light-chain variants in different IGLV families with and without V<sub>L</sub> 91<sub>F/Y/W</sub>/96<sub>non-F/Y/W</sub> are compared. Red: with V<sub>L</sub> 91<sub>F/Y/W</sub>/96<sub>non-F/Y/W</sub>; blue: without V<sub>L</sub> 91<sub>F/Y/W</sub>/96<sub>non-F/Y/W</sub>. P values were computed by two-tailed Student's t test.

P values were computed by two-tailed Student's t test.

See also Figures S2 and S3, and Table S2.

of light-chain variants with and without V<sub>L</sub> 91<sub>F/Y/W</sub>/96<sub>non-F/Y/W</sub> had no significant difference ( $p = 0.31$ – $0.84$ , two-tailed Student's t test; Figure S3A). These results show that besides the amino acid sequences at V<sub>L</sub> residues 91 and 96, sequence differences among IGLV families could also influence the binding activity of CR9114 to the HA stem.

Our additional analysis suggested that a CDR L3 length of 11 was optimal for CR9114 binding to the HA stem (Figure S3B) since the binding scores of light-chain variants lowered when the CDR L3 lengths deviated from 11. In contrast, the expression scores were similar across light-chain variants with different

CDR L3 lengths (Figure S3C). We also analyzed the relationship between binding scores and the number of V<sub>L</sub> SHMs, but no significant correlation was found ( $p = 0.39$ – $0.98$ , Pearson correlation test; Figures S3D–S3F). Overall, our analyses demonstrate that light-chain germline features, including CDR L3 length and germline gene usage, are important sequence determinants for the HA-stem-binding activity of CR9114.

### High-throughput characterization of CDR H3 variants

Similar to the light chain, the sequence constraints of the CDR H3 of IGHV1-69 HA stem antibodies have also been unclear, especially since they are highly diverse.<sup>14</sup> Therefore, we next aimed to probe the sequence constraints of CR9114 CDR H3. Briefly, we compiled a list of 3,325 CDR H3 sequences from diverse IGHV1-69 antibodies from GenBank (Table S3). Additionally, we performed a site-saturation mutagenesis of the CDR H3 of CR9114. Together with the WT CDR H3 sequence of CR9114, a CDR H3 library with 3,606 variants in CR9114 gHC was constructed and displayed on the yeast cell surface. Tite-Seq<sup>18</sup> was then applied to measure the apparent dissociation constant ( $K_D$ ) values of individual variants to mini-HA (Figure 3A). At the same time, antibody expression level was measured by three-way cell sorting and next-generation sequencing, as described above for the light-chain variant library. The expression score for each CDR H3 variant was normalized such that the score for WT equaled 1 and the mean score for nonsense mutants of CR9114 CDR H3 equaled 0.

After filtering out CDR H3 variants with low occurrence frequency or noisy estimation in the apparent  $K_D$  values (see STAR Methods), 2,162 CDR H3 variants were subjected to downstream analyses (Table S3). These included the WT CDR H3 of CR9114, 206 single amino acid mutants of CR9114 CDR H3, 14 nonsense mutants of CR9114 CDR H3, 73 CDR H3 variants from HA stem antibodies, 245 from non-HA stem influenza antibodies, and 1,623 from non-influenza antibodies. Pearson correlations of 0.71 and 0.61 were obtained between two replicates of Tite-Seq and expression sort, respectively (Figures S4A–S4B), confirming the reproducibility of our results. As expected, the expression score of nonsense mutants of CR9114 CDR H3 was significantly lower than other CDR H3 variants ( $p = 2e-22$  to  $1e-47$ , two-tailed Student's t test; Figure S4D). Similarly, the binding affinity of nonsense mutants of CR9114 CDR H3 to mini-HA was significantly weaker than that of single amino acid mutants of CR9114 CDR H3 ( $p = 1e-17$ , two-tailed Student's t test; Figures 3B and 3C). These results validated the quality of our data.

### Most CDR H3 variants are incompatible with CR9114 for HA stem binding

While the apparent  $K_D$  values of many single amino acid mutants of CR9114 CDR H3 to mini-HA were around 1 nM, those of CDR H3 variants from other antibodies, including HA stem antibodies, were mostly between 100 nM and 1  $\mu$ M (Figures 3B and 3C). In contrast, the expression scores of single amino acid mutants of CR9114 CDR H3 and other CDR H3 variants were similar ( $p = 0.22$ – $0.76$ , two-tailed Student's t test; Figure S4D). Of note, several CDR H3 variants from non-influenza antibodies had an apparent  $K_D$  of around 1 nM (Figure S4C; Table S3). However, when we recombinantly expressed one of these CDR H3 variants,

**Table 1. Binding affinity of CR9114 gHC mutants against mini-HA and SI06 HA**

CR9114 gHC mutant Fab	$K_D$ ( $\mu$ M)	
	vs. mini-HA	vs. SI06 HA
WT	6.2	2.6
V <sub>L</sub> W91Y	45.7	27.2
V <sub>L</sub> W91F	9.7	9.6
V <sub>L</sub> W91T	N.B.	N.B.
V <sub>L</sub> W91R	N.B.	N.B.
V <sub>L</sub> W91A	N.B.	N.B.
V <sub>L</sub> A96Y	N.B.	N.B.
V <sub>L</sub> A96W	N.B.	N.B.
V <sub>L</sub> A96F	N.B.	N.B.
V <sub>L</sub> A96V	3.8	15.8
V <sub>L</sub> A96S	1.9	1.6
V <sub>L</sub> A96R	13.0	21.5

Related to Figure S2. N.B., no binding.

it did not show binding to mini-HA, indicating there were false positives in our Tite-Seq experiment (Figure S4E). Together, these observations suggest that the CDR H3 of CR9114 has a stringent sequence requirement for HA stem binding.

A previous study has demonstrated that many IGHV1-69 HA stem antibodies, including CR9114, are featured by a CDR H3-encoded Tyr at V<sub>H</sub> residue 98 that interacts extensively with the HA stem epitope.<sup>14</sup> Consistently, our results showed that CDR H3 variants with Y98 had slightly yet significantly better apparent  $K_D$  values ( $p = 0.001$ – $0.03$ , two-tailed Student's *t* test; Figure 3D). In contrast, CDR H3 variants with and without Tyr, regardless of the residue position, did not show any significant difference in apparent  $K_D$  values ( $p = 0.20$ – $0.91$ , two-tailed Student's *t* test; Figure 3E). Therefore, our result substantiates that V<sub>H</sub> Y98 partially contributes to the compatibility of CDR H3 variants with CR9114 for HA stem binding. However, given that many CDR H3 variants from IGHV1-69 HA stem antibodies with Y98 are incompatible with CR9114 gHC for binding (Figure 3D), the sequence constraints of CR9114 CDR H3 likely involve other CDR H3 residues.

### Importance of non-paratope residues in CR9114 CDR H3

To further understand the sequence constraints of CR9114 CDR H3, we aimed to identify residues that are key for HA stem binding. Subsequently, we analyzed the single amino acid mutants of CR9114 CDR H3 in our CDR H3 library. Our Tite-Seq result indicated that most mutations at the four consecutive Tyr residues in the CDR H3, namely V<sub>H</sub> Y98, V<sub>H</sub> Y99, V<sub>H</sub> Y100, and V<sub>H</sub> Y100a, weakened the binding affinity of CR9114 gHC (Figure 4A). This observation is consistent with our structural analysis above (Figure 2A) showing that these four Tyr residues form an extensive  $\pi$ - $\pi$  stacking network with V<sub>H</sub> W47 and V<sub>L</sub> W91, which is essential for the conformational stability of the CDR H3. Additionally, our Tite-Seq result also revealed the low mutational tolerance of V<sub>H</sub> H95 and V<sub>H</sub> N97, hence their importance in the HA-stem-binding activity of CR9114 (Figure 4A).

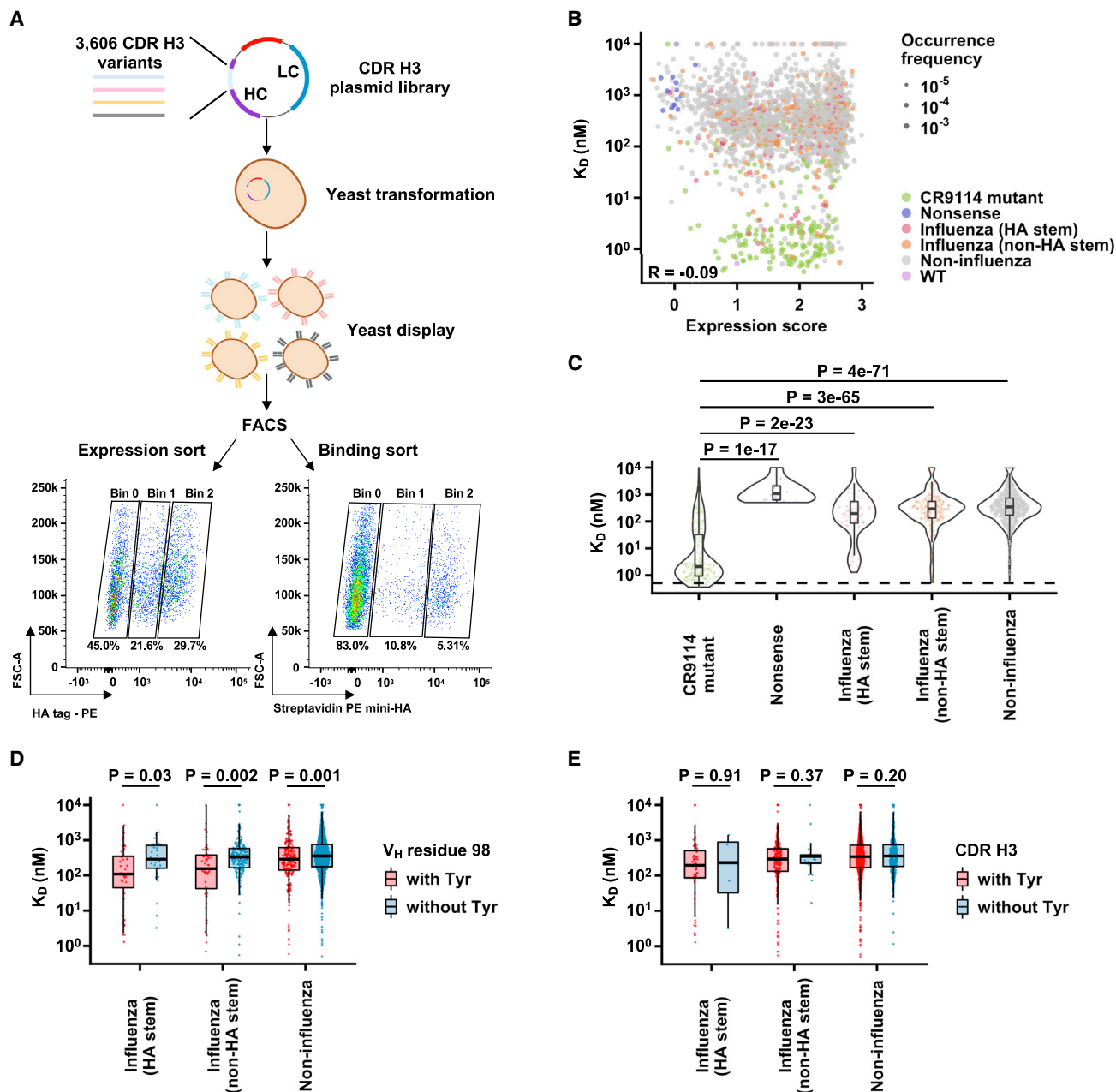
Based on a previously determined crystal structure of CR9114 in complex with HA,<sup>7</sup> the side chains of both V<sub>H</sub> H95 and V<sub>H</sub> N97 are not interacting with the HA stem epitope. Instead, both V<sub>H</sub> H95 and V<sub>H</sub> N97 form intramolecular interactions to stabilize the CDR H3 conformation. V<sub>H</sub> H95 H-bonds with V<sub>H</sub> S100b and V<sub>H</sub> S35 as well as interacts with V<sub>H</sub> Y100 via T-shaped  $\pi$ - $\pi$  stacking (Figure 4B), whereas V<sub>H</sub> N97 H-bonds with V<sub>H</sub> Y99 and V<sub>H</sub> S100b (Figure 4C). These observations corroborate our light-chain analysis, demonstrating that non-paratope residues that stabilize the CDR H3 conformation are important for the HA-stem-binding activity of CR9114.

### DISCUSSION

IGHV1-69 is one of the most highly used heavy-chain V genes in the human antibody repertoire,<sup>19</sup> suggesting its importance in the immune system. In fact, many known broadly neutralizing antibodies to various pathogens, such as influenza virus, hepatitis C virus (HCV), and human immunodeficiency virus (HIV), were encoded by IGHV1-69.<sup>15</sup> As a result, IGHV1-69 is regarded as an S.O.S. component of the human antibody repertoire.<sup>20</sup> However, sequence determinants for the antigen specificity of IGHV1-69 antibodies have been largely elusive. In this study, we used a high-throughput approach to probe the sequence determinants for the HA-stem-binding activity of CR9114, an IGHV1-69 broadly neutralizing antibody to influenza HA stem.<sup>7</sup> Our results revealed the importance of the CR9114 light chain in binding, despite having no contact with the epitope. In addition, we showed that the CDR H3 sequence of CR9114 has stringent sequence constraints. Overall, this work advances our understanding of the sequence determinants that define IGHV1-69 HA stem antibodies.

Our results show that V<sub>L</sub> residue 96 in CDR L3 is a determinant for the HA-stem-binding activity of CR9114. V<sub>L</sub> residue 96 in some IGHV1-69 antibodies, including CR9114, is encoded by the light-chain J gene (Figure S5A). Among seven known IGLJ germline genes, three (IGLJ1, IGLJ4, and IGLJ5) encode an aromatic amino acid at V<sub>L</sub> residue 96 (Figure S5B). Given that aromatic amino acids are forbidden at V<sub>L</sub> residue 96 of CR9114, these observations suggest that the light-chain J gene plays a role in generating IGHV1-69 HA stem antibodies. Of note, the contribution of the light-chain J gene to antibody binding is rarely reported, if at all, since most antibody studies focus on the V genes. However, the sequence determinants for the binding activity of IGHV1-69 HA stem antibodies likely vary from antibody to antibody (see discussion below). As a result, additional work is needed to decipher whether light-chain J gene usage is a common sequence constraint of IGHV1-69 HA stem antibodies.

Perhaps the most perplexing result in our study is that most CDR H3 sequences from other IGHV1-69 HA stem antibodies are incompatible with CR9114. Provided that diverse light-chain sequences can be observed in IGHV1-69 HA stem antibodies,<sup>7,9,11–13</sup> the compatibility of a given CDR H3 sequence in IGHV1-69 HA stem antibodies likely depends on their light-chain sequences. In other words, we postulate that whether an IGHV1-69 antibody can bind to HA stem is coordinately determined by the light-chain and CDR H3 sequences. Consistently, our results revealed the importance of light chain-CDR H3



**Figure 3. Yeast display of CDR H3 library of CR9114 gHC**

(A) Schematic illustration of measuring the expression level and HA-stem-binding activity of many CDR H3 variants in parallel using yeast surface display. Briefly, a CDR H3 library of CR9114 gHC was displayed on the yeast cell surface and subjected to FACS based on surface expression level and binding to mini-HA. The sorted populations were analyzed by next-generation sequencing. Two independent replicates were performed.

(B) Relationship between apparent dissociation constant ( $K_D$ ) values and expression scores. Greenish yellow: CR9114 single amino acid mutants; purple: nonsense variants with stop codons; red: CDR H3 variants from IGHV1-69 HA stem antibodies; orange: CDR H3 variants from IGHV1-69 non-HA stem influenza antibodies; gray: CDR H3 variants from IGHV1-69 non-influenza antibodies; pink: WT. The Pearson correlation coefficient ( $R$ ) is indicated.

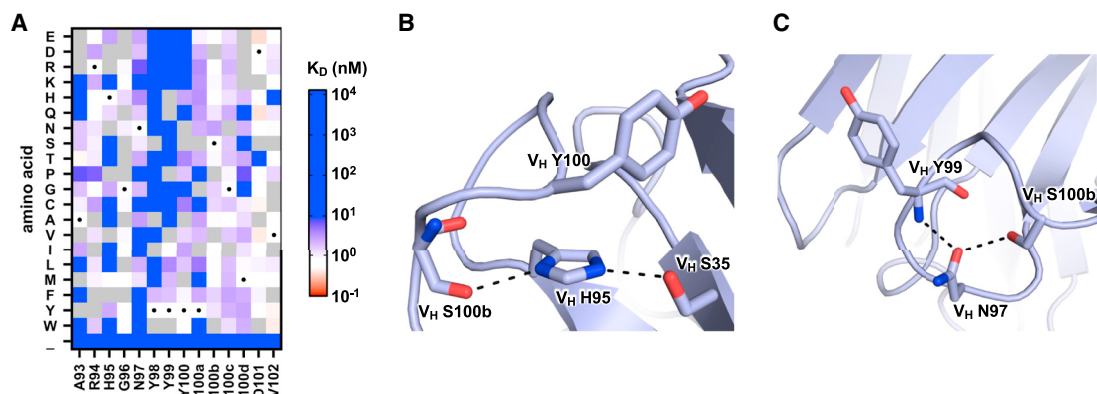
(C) Distribution of the apparent  $K_D$  values among CDR H3 variants from different types of antibodies.

(D) Comparison of apparent  $K_D$  values between antibodies with and without V<sub>H</sub> Y98.

(E) Comparison of apparent  $K_D$  values between antibodies with and without a Tyr in their CDR H3.

P values were computed by two-tailed Student's t test.

See also Figure S4 and Table S3.



**Figure 4. Importance of V<sub>H</sub> H95 and N97 for HA-stem-binding activity of CR9114**

(A) Apparent  $K_D$  values of individual mutations in the CDR H3 of CR9114 are shown as a heatmap. WT amino acids are indicated by a black circle. Mutations in gray were excluded in our data analysis due to low input count or high occurrence in the WT sample (see STAR Methods).

(B) Intramolecular interactions involving CR9114 V<sub>H</sub> H95 are shown.

(C) Intramolecular interactions involving CR9114 V<sub>H</sub> N97 are shown. Hydrogen bonds are represented by dashed lines. PDB: 4FQI is used.<sup>7</sup>

See also Figure S5.

interaction in the HA-stem-binding activity of CR9114. Akin to CR9114, both CR6261 and F10, which are two other IGHV1-69 HA stem antibodies,<sup>11,12</sup> feature an aromatic Trp at V<sub>L</sub> residue 91 and a small amino acid Val at V<sub>L</sub> residue 96, positioning toward the heavy chain with limited space in between (Figures S5C–S5D). Furthermore, V<sub>L</sub> W91 of both CR6261 and F10 is also involved in the  $\pi$ - $\pi$  stacking interaction with the CDR H3 (Figures S5E–S5F). Therefore, the sequence determinants of IGHV1-69 HA stem antibodies are stringent yet complex, which explain the limited sequence convergence among IGHV1-69 HA stem antibodies. Comprehending these sequence determinants will enable an accurate estimation of the proportion of B cells that can give rise to IGHV1-69 HA stem antibodies, especially those that can cross-react to both group 1 and 2 HAs as well as influenza B HA, like CR9114. As the efforts to develop a universal influenza vaccine continue,<sup>21–23</sup> future studies on the sequence determinants of IGHV1-69 HA stem antibodies are warranted.

### Limitations of the study

Since our approach has focused on the  $\lambda$  light chain of IGHV1-69 antibodies, we were unable to demonstrate whether the  $\kappa$  light chain has similar sequence constraints when paired with CR9114 heavy chain for HA stem binding. Another limitation of our study is that only five antigen concentrations were used in our Tite-Seq experiment as opposed to the >10 used in other Tite-Seq studies.<sup>18,24,25</sup> This shortcoming would increase the estimation error of apparent  $K_D$  values.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

- Lead contact
- Materials availability
- Data and code availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Cell lines
- METHOD DETAILS
  - CR9114 gHC yeast display plasmid
  - Construction of CDR H3 library and light chain variant library
  - Yeast transformation
  - Expression and purification of mini-HA and SI06 HA
  - Biotinylation and PE-conjugation of mini-HA
  - Fluorescence-activated cell sorting (FACS) of yeast display library
  - Next-generation sequencing of light chain variant library and CDR H3 library
  - Computing the binding score, expression score, and apparent  $K_D$  values
  - Expression and purification of Fabs
  - Biolayer interferometry binding assay
- QUANTIFICATION AND STATISTICAL ANALYSIS

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2023.113410>.

### ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health R01 AI167910 (N.C.W.), DP2 AT011966 (N.C.W.), the Searle Scholars Program (N.C.W.), and Howard Hughes Medical Institute Emerging Pathogens Initiative (N.C.W.).

### AUTHOR CONTRIBUTIONS

N.C.W. conceived and designed the study. Y.W., H.L., and N.C.W. assembled the dataset. Q.W.T. constructed the yeast libraries and prepared the sequencing libraries. Y.W. and N.C.W. performed data analysis. H.L.,



T.J.C.T., R.L., and K.J.M. assisted with experiments. Q.W.T., Y.W., H.L., and N.C.W. wrote the paper, and all authors reviewed and/or edited the paper.

## DECLARATION OF INTERESTS

N.C.W. consults for HeliXon.

Received: July 27, 2023

Revised: September 29, 2023

Accepted: October 24, 2023

Published: November 16, 2023

## REFERENCES

- Petrova, V.N., and Russell, C.A. (2018). The evolution of seasonal influenza viruses. *Nat. Rev. Microbiol.* **16**, 47–60.
- Frentzel, E., Jump, R.L.P., Archbald-Pannone, L., Nace, D.A., Schweon, S.J., Gaur, S., Naqvi, F., Pandya, N., and Mercer, W.; Infection Advisory Subcommittee of AMDA, The Society for Post-Acute and Long-Term Care Medicine (2020). Recommendations for mandatory influenza vaccinations for health care personnel from AMDA's Infection Advisory Subcommittee. *J. Am. Med. Dir. Assoc.* **21**, 25–28.e2.
- Tricco, A.C., Chit, A., Soobiah, C., Hallett, D., Meier, G., Chen, M.H., Tashkandi, M., Bauch, C.T., and Loeb, M. (2013). Comparing influenza vaccine efficacy against mismatched and matched strains: a systematic review and meta-analysis. *BMC Med.* **11**, 153–219.
- Carrat, F., and Flahault, A. (2007). Influenza vaccine: the challenge of antigenic drift. *Vaccine* **25**, 6852–6862.
- Impagliazzo, A., Milder, F., Kuipers, H., Wagner, M.V., Zhu, X., Hoffman, R.M.B., van Meersbergen, R., Huizingh, J., Wanningen, P., Verspuij, J., et al. (2015). A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen. *Science* **349**, 1301–1306.
- Yassine, H.M., Boyington, J.C., McTamney, P.M., Wei, C.-J., Kanekiyo, M., Kong, W.-P., Gallagher, J.R., Wang, L., Zhang, Y., Joyce, M.G., et al. (2015). Hemagglutinin-stem nanoparticles generate heterosubtypic influenza protection. *Nat. Med.* **21**, 1065–1070.
- Dreyfus, C., Laursen, N.S., Kwaks, T., Zuidgeest, D., Khayat, R., Ekiert, D.C., Lee, J.H., Metlagel, Z., Bujny, M.V., Jongeneelen, M., et al. (2012). Highly conserved protective epitopes on influenza B viruses. *Science* **337**, 1343–1348.
- Andrews, S.F., Graham, B.S., Mascola, J.R., and McDermott, A.B. (2018). Is it possible to develop a “universal” influenza virus vaccine? Immunogenetic considerations underlying B-cell biology in the development of a pan-subtype influenza A vaccine targeting the hemagglutinin stem. *Cold Spring Harbor Perspect. Biol.* **10**, a029413.
- Throsby, M., van den Brink, E., Jongeneelen, M., Poon, L.L.M., Alard, P., Cornelissen, L., Bakker, A., Cox, F., van Deventer, E., Guan, Y., et al. (2008). Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells. *PLoS One* **3**, e3942.
- Pappas, L., Foglierini, M., Piccoli, L., Kallewaard, N.L., Turrini, F., Silacci, C., Fernandez-Rodriguez, B., Agatic, G., Giacchetto-Sasselli, I., Pellicciotta, G., et al. (2014). Rapid development of broadly influenza neutralizing antibodies through redundant mutations. *Nature* **516**, 418–422.
- Sui, J., Hwang, W.C., Perez, S., Wei, G., Aird, D., Chen, L.-m., Santelli, E., Stec, B., Cadwell, G., Ali, M., et al. (2009). Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat. Struct. Mol. Biol.* **16**, 265–273.
- Ekiert, D.C., Bhabha, G., Elsliger, M.-A., Friesen, R.H.E., Jongeneelen, M., Throsby, M., Goudsmit, J., and Wilson, I.A. (2009). Antibody recognition of a highly conserved influenza virus epitope. *Science* **324**, 246–251.
- Lang, S., Xie, J., Zhu, X., Wu, N.C., Lerner, R.A., and Wilson, I.A. (2017). Antibody 27F3 broadly targets influenza A group 1 and 2 hemagglutinins through a further variation in V<sub>H</sub>1-69 antibody orientation on the HA stem. *Cell Rep.* **20**, 2935–2943.
- Avnir, Y., Tallarico, A.S., Zhu, Q., Bennett, A.S., Connelly, G., Sheehan, J., Sui, J., Fahmy, A., Huang, C.-y., Cadwell, G., et al. (2014). Molecular signatures of hemagglutinin stem-directed heterosubtypic human neutralizing antibodies against influenza A viruses. *PLoS Pathog.* **10**, e1004103.
- Chen, F., Tzarum, N., Wilson, I.A., and Law, M. (2019). V<sub>H</sub>1-69 antiviral broadly neutralizing antibodies: genetics, structures, and relevance to rational vaccine design. *Curr. Opin. Virol.* **34**, 149–159.
- Xu, R., Krause, J.C., McBride, R., Paulson, J.C., Crowe, J.E., Jr., and Wilson, I.A. (2013). A recurring motif for antibody recognition of the receptor-binding site of influenza hemagglutinin. *Nat. Struct. Mol. Biol.* **20**, 363–370.
- Lee, P.S., Ohshima, N., Stanfield, R.L., Yu, W., Iba, Y., Okuno, Y., Kurosawa, Y., and Wilson, I.A. (2014). Receptor mimicry by antibody F045–092 facilitates universal binding to the H3 subtype of influenza virus. *Nat. Commun.* **5**, 3614.
- Phillips, A.M., Lawrence, K.R., Moulana, A., Dupic, T., Chang, J., Johnson, M.S., Cvijovic, I., Mora, T., Walczak, A.M., and Desai, M.M. (2021). Binding affinity landscapes constrain the evolution of broadly neutralizing anti-influenza antibodies. *Elife* **10**, e71393.
- Boyd, S.D., Gaëta, B.A., Jackson, K.J., Fire, A.Z., Marshall, E.L., Merker, J.D., Maniar, J.M., Zhang, L.N., Sahaf, B., Jones, C.D., et al. (2010). Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J. Immunol.* **184**, 6986–6992.
- Lerner, R.A. (2011). Rare antibodies from combinatorial libraries suggests an SOS component of the human immunological repertoire. *Mol. Biosyst.* **7**, 1004–1012.
- Erbelding, E.J., Post, D.J., Stemmy, E.J., Roberts, P.C., Augustine, A.D., Ferguson, S., Paules, C.I., Graham, B.S., and Fauci, A.S. (2018). A universal influenza vaccine: the strategic plan for the National Institute of Allergy and Infectious Diseases. *J. Infect. Dis.* **218**, 347–354.
- Nachbagauer, R., Feser, J., Naficy, A., Bernstein, D.I., Guptill, J., Walter, E.B., Berlanda-Scorza, F., Stadlbauer, D., Wilson, P.C., Aydiello, T., et al. (2021). A chimeric hemagglutinin-based universal influenza virus vaccine approach induces broad and long-lasting immunity in a randomized, placebo-controlled phase I trial. *Nat. Med.* **27**, 106–114.
- Andrews, S.F., Cominsky, L.Y., Shimberg, G.D., Gillespie, R.A., Gorman, J., Raab, J.E., Brand, J., Creanga, A., Gajjala, S.R., Narpala, S., et al. (2023). An influenza H1 hemagglutinin stem-only immunogen elicits a broadly cross-reactive B cell response in humans. *Sci. Transl. Med.* **15**, eade4976.
- Adams, R.M., Mora, T., Walczak, A.M., and Kinney, J.B. (2016). Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *Elife* **5**, e23156.
- Starr, T.N., Greaney, A.J., Hilton, S.K., Ellis, D., Crawford, K.H.D., Diggins, A.S., Navarro, M.J., Bowen, J.E., Tortorici, M.A., Walls, A.C., et al. (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310.e20.
- Chao, G., Lau, W.L., Hackel, B.J., Sazinsky, S.L., Lippow, S.M., and Wittrup, K.D. (2006). Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**, 755–768.
- Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013). Nucleic Acids Res. **41**, D36–D42.
- Ye, J., Ma, N., Madden, T.L., and Ostell, J.M. (2013). IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–W40.
- Benatui, L., Perez, J.M., Belk, J., and Hsieh, C.-M. (2010). An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155–159.
- Ekiert, D.C., Friesen, R.H.E., Bhabha, G., Kwaks, T., Jongeneelen, M., Yu, W., Ophorst, C., Cox, F., Korse, H.J.W.M., Brandenburg, B., et al. (2011). A highly conserved neutralizing epitope on group 2 influenza A viruses. *Science* **333**, 843–850.

31. Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., and Kanitz, A. (2021). Sustainable data analysis with Snakemake. *F1000Research* 10.
32. Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614–620.
33. Tareen, A., and Kinney, J.B. (2020). Logomaker: beautiful sequence logos in Python. *Bioinformatics* 36, 2272–2274.
34. Wu, N.C., Grande, G., Turner, H.L., Ward, A.B., Xie, J., Lerner, R.A., and Wilson, I.A. (2017). In vitro evolution of an influenza broadly neutralizing antibody is modulated by hemagglutinin receptor specificity. *Nat. Commun.* 8, 15371.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
PE anti-HA.11	Biolegend	Cat# 901517; RRID: AB_2629623
<b>Chemicals, peptides, and recombinant proteins</b>		
Acid Casein Peptone (Casamino Acids)	Fisher Scientific	Cat# BP1424-100
D-Galactose	Fisher Scientific	Cat# BP656-500
D-Glucose	Fisher Scientific	Cat# D16-500
D-(+)-Raffinose Pentahydrate	Thermo Fisher Scientific	Cat# J21060-36
Dithiothreitol	Thermo Fisher Scientific	Cat# A39255
Lithium Acetate Dihydrate	Sigma-Aldrich	Cat# L6883-250G
Sorbitol	Fisher Scientific	Cat# BP-439-500
Sodium Phosphate Monobasic Monohydrate (NaH <sub>2</sub> PO <sub>4</sub> · H <sub>2</sub> O)	VWR	Cat# BDH9298-500G
Sodium Phosphate Dibasic Anhydrous (Na <sub>2</sub> HPO <sub>4</sub> )	Fisher Scientific	Cat# BP332-500
Streptavidin-PE Conjugate	Invitrogen	Cat# S866
Yeast Nitrogen Base with Ammonium Sulfate	MP Biomedicals	Cat# 4027512
Sodium chloride (NaCl)	Fisher Scientific	Cat# S271-500
Hydrochloric Acid, ACS, 12M	Fisher Scientific	Cat# S25358
Tris Base	Fisher Scientific	Cat# BP152-500
Tris-HCl	Fisher Scientific	Cat# BP153-500
Imidazole	Fisher Scientific	Cat# A10221
Calcium Chloride, Dihydrate	Millipore	Cat# 208291-250GM
DpnI	New England Biolabs	Cat# R0176L
Cellfectin II Reagent	Gibco	Cat# 10362-100
Sf-900 II SFM	Thermo Fisher Scientific	Cat# 10902088
Ni Sepharose Excel Resin	Cytiva	Cat# 17371202
XK 16/100 Superdex 200 pg	Cytiva	Cat# 90100137
Phosphate-buffered saline (PBS), 1X	VWR	Cat# 21-040-CM
BSA	Sigma-Aldrich	Cat# A9418
Tween 20	Fisher Scientific	Cat# BP337-500
NEB 5-alpha Competent E.coli	New England Biolabs	Cat# C2987H
MAX Efficiency DH10Bac Competent Cells	Thermo Fisher Scientific	Cat# 10361012
<b>Critical commercial assays</b>		
Biotin-Protein Ligase-BIRA kit	Avidity	Cat# BIRA 500
Expi293 Expression System Kit	Thermo Fisher Scientific	Cat# A14635
Monarch DNA Gel Extraction Kit	New England Biolabs	Cat# T1020L
PureLink PCR Purification Kit	Thermo Fisher Scientific	Cat# K210003
KOD Hot Start DNA Polymerase	EMD Millipore	Cat# 71086-3
PrimeSTAR MAX DNA Polymerase	Takara	Cat# R045A
QuikChange XL Site-Directed Mutagenesis kit	Agilent	Cat# 200516
Q5 High-Fidelity DNA polymerase	New England Biolabs	Cat# M0491L
Octet HIS1K biosensors	Sartorius	Cat# 18-5120
Zymoprep Yeast Plasmid Miniprep II Kit	Zymo Research	Cat# D2004
PureLink HiPure Plasmid Miniprep Kit	Thermo Fisher Scientific	Cat# K210003

(Continued on next page)

<b>Continued</b>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw sequencing reads	This study	BioProject: PRJNA976657
Experimental models: Cell lines		
Expi293F cells	Thermo Fisher Scientific	Cat# A14527
<i>Saccharomyces cerevisiae</i> EBY100 cells	ATCC	Cat# MYA-4941
Sf9 cells	ATCC	Cat# CRL-1711
Recombinant DNA		
phCMV3_CR9114_Fab	This study	N/A
pCTcon2_CR9114_GL	This study	N/A
pFastBac-mini-HA	This study	N/A
pFastBac-SI06-HA	This study	N/A
Software and algorithms		
FlowJo v10.8	BD Life Sciences	N/A
Octet analysis software 9.0	Sartorius	N/A
Python	<a href="https://www.python.org/">https://www.python.org/</a>	N/A
R	<a href="https://www.r-project.org/">https://www.r-project.org/</a>	N/A
Custom scripts	This study	<a href="https://doi.org/10.5281/zenodo.8382109">https://doi.org/10.5281/zenodo.8382109</a>

## RESOURCE AVAILABILITY

### Lead contact

Information and requests for resources should be directed and will be fulfilled by the lead contact, Nicholas C. Wu ([nicwu@illinois.edu](mailto:nicwu@illinois.edu)).

### Materials availability

All plasmids generated in this study are available from the [lead contact](#) without restriction.

### Data and code availability

- Raw sequencing data have been submitted to the NIH Short Read Archive under BioProject: PRJNA976657.
- Custom python scripts for all analyses have been deposited to <https://doi.org/10.5281/zenodo.8382109>.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODELS AND SUBJECT DETAILS

### Cell lines

*Saccharomyces cerevisiae* EBY100 cells (American Type Culture Collection) was culture at 30°C (except where indicated) in baffled flasks while shaking at 225 rpm. Selection media (SD-CAA) contained 2% w/v D-glucose, 0.67% w/v yeast nitrogen base with ammonium sulfate, 0.5% w/v casamino acids, 0.54% w/v Na<sub>2</sub>HPO<sub>4</sub>, 0.86% w/v NaH<sub>2</sub>PO<sub>4</sub> · H<sub>2</sub>O, all dissolved in deionized water. Induction media (SGR-CAA) contained 2% w/v galactose, 2% w/v raffinose, 0.1% w/v D-glucose, 0.67% w/v yeast nitrogen base with ammonium sulfate, 0.5% w/v casamino acids, 0.54% w/v Na<sub>2</sub>HPO<sub>4</sub>, 0.86% w/v NaH<sub>2</sub>PO<sub>4</sub> · H<sub>2</sub>O, all dissolved in deionized water. Expi293F (Thermo Fisher Cat no. A14527) suspension cells were grown at 37°C in a humidified 8% CO<sub>2</sub> incubator rotating at 130rpm. Sf9 (ATCC Cat no. CRL-1711) were maintained in Sf-900 SFM medium (Thermo Fisher Scientific).

## METHOD DETAILS

### CR9114 gHC yeast display plasmid

CR9114 gHC yeast display plasmid, pCTcon2\_CR9114\_GL, was generated by cloning the coding sequences of (from N-terminal to C-terminal, all in-frame) Aga2 secretion signal, CR9114 wild-type Fab light chain, V5 tag, equine rhinitis B virus (ERBV-1) 2A self-cleaving peptide, Aga2 secretion signal, CR9114 germline Fab heavy chain, HA tag, and Aga2p, into the pCTcon2 vector.<sup>26</sup>



### Construction of CDR H3 library and light chain variant library

Sequences of IGHV1-69 antibodies were obtained from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>).<sup>27</sup> IgBLAST was used to identify the CDR H3 region.<sup>28</sup> Light chain variant library (Table S1) and CDR H3 library (Table S3) were synthesized as oligo pools by Integrated DNA Technologies and Twist Bioscience, respectively. Names and sequences of primers for cloning the libraries into pCTcon2\_CR9114\_GL are listed in Table S4. Oligo pool of the light chain variant library was PCR-amplified using primers IGHV1-69-Lightchain-lib-IF and IGHV1-69-Lightchain-lib-IR. Then, the amplified oligonucleotide pool was gel-purified using a Monarch DNA Gel Extraction Kit (NEB). To generate the linearized vector for the light chain variant library, pCTcon2\_CR9114\_GL was used as a template for PCR using primers IGHV1-69-Lightchain-lib-VF and IGHV1-69-Lightchain-lib-VR. The PCR product was then gel-purified. Similarly, oligo pool of the CDR H3 library was PCR-amplified using primers IGHV1-69-CDRH3-lib-IF and IGHV1-69-CDRH3-lib-IR. Then, the amplified oligo pool was gel-purified using a Monarch DNA Gel Extraction Kit (NEB). To generate the linearized vector for the light chain variant library, pCTcon2\_CR9114\_GL was used as a template for PCR using primers IGHV1-69-CDRH3-lib-VF and IGHV1-69-CDRH3-lib-VR. The PCR product was then gel-purified. All PCRs were performed using KOD Hot Start DNA polymerase (EMD Millipore) according to the manufacturer's instructions.

### Yeast transformation

Yeast cells were transformed by electroporation following a previously described protocol.<sup>29</sup> Briefly, *Saccharomyces cerevisiae* EBY100 cells (American Type Culture Collection) were grown in YPD medium (1% w/v yeast nitrogen base, 2% w/v peptone, 2% w/v D(+)-glucose) overnight at 30°C with shaking at 225 rpm until OD<sub>600</sub> reached 3. Then, an aliquot of overnight culture was grown in 100 mL YPD media, with an initial OD<sub>600</sub> of 0.3, shaking at 225 rpm at 30°C. Once OD<sub>600</sub> reached 1.6, yeast cells were collected by centrifugation at 1700 × g for 3 min at room temperature. Media were removed and the cell pellet was washed twice with 50 mL ice-cold water, and then once with 50 mL of ice-cold electroporation buffer (1 M sorbitol, 1 mM calcium chloride). Cells were resuspended in 20 mL conditioning media (0.1 M lithium acetate, 10 mM dithiothreitol) and shaken at 225 rpm at 30°C. Cells were collected via centrifugation at 1700 × g for 3 min at room temperature, washed once with 50 mL ice-cold electroporation buffer, resuspended in electroporation buffer to reach a final volume of 1 mL, and kept on ice. 5 µg of the amplified oligo pool (light chain variant library or CDR H3 library) and 4 µg of the corresponding purified linearized vector were added into 400 µL of conditioned yeast. The mixture was transferred to a pre-chilled BioRad GenePulser cuvette with 2 mm electrode gap and kept on ice for 5 min until electroporation. Cells were electroporated at 2.5 kV and 25 µF, achieving a time constant between 3.7 and 4.1 ms. Electroporated cells were transferred into 4 mL of YPD media supplemented with 4 mL of 1 M sorbitol and incubated at 30°C with shaking at 225 rpm for 1 h. Cells were collected via centrifugation at 1700 × g for 3 min at room temperature, resuspended in 0.6 mL SD-CAA medium (2% w/v D-glucose, 0.67% w/v yeast nitrogen base with ammonium sulfate, 0.5% w/v casamino acids, 0.54% w/v Na<sub>2</sub>HPO<sub>4</sub>, 0.86% w/v NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O, all dissolved in deionized water), plated onto SD-CAA plates (2% w/v D-glucose, 0.67% w/v yeast nitrogen base with ammonium sulfate, 0.5% w/v casamino acids, 0.54% w/v Na<sub>2</sub>HPO<sub>4</sub>, 0.86% w/v NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O, 18.2% w/v sorbitol, 1.5% w/v agar, all dissolved in deionized water) and incubated at 30°C for 40 h. Colonies were then collected in SD-CAA medium, centrifuged at 1700 × g for 5 min at room temperature, and resuspended in SD-CAA medium with 15% v/v glycerol such that OD<sub>600</sub> was 50. Glycerol stocks were stored at −80°C until used.

### Expression and purification of mini-HA and SI06 HA

Oligonucleotide encoding mini-HA #4900 protein<sup>5</sup> or SI06 HA was fused with N-terminal gp67 signal peptide and a C-terminal BirA biotinylation site, thrombin cleavage site, trimerization domain, and a His<sub>6</sub> tag, and then cloned into a customized baculovirus transfer vector.<sup>30</sup> Recombinant bacmid DNA that carried the protein construct was generated using the Bac-to-Bac system (Thermo Fisher Scientific) according to the manufacturer's instructions. Baculovirus was generated by transfecting the purified bacmid DNA into adherent Sf9 cells using Cellfectin reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. The baculovirus was further amplified by passaging in adherent Sf9 cells at a multiplicity of infection (MOI) of 1. Recombinant protein was expressed by infecting 1 L of suspension Sf9 cells at an MOI of 1. On day 3 post-infection, Sf9 cells were pelleted by centrifugation at 4000 × g for 25 min, and soluble recombinant protein was purified from the supernatant by affinity chromatography using Ni Sepharose excel resin (Cytiva) and then size exclusion chromatography using a HiLoad 16/100 Superdex 200 prep grade column (Cytiva) in 20 mM Tris-HCl pH 8.0, 100 mM NaCl. The purified protein was concentrated by Amicon spin filter (Millipore Sigma) and filtered by 0.22 µm centrifuge Tube Filters (Costar). Concentration of the protein was determined by nanodrop (Fisher Scientific). Protein was subsequently aliquoted, flash frozen in dry-ice ethanol mixture, and store at −80°C until used.

### Biotinylation and PE-conjugation of mini-HA

Purified mini-HA was biotinylated using the Biotin-Protein Ligase-BIRA kit according to the manufacturer's instructions (Avidity). Biotinylated mini-HA was then conjugated to streptavidin-PE (Thermo Fisher Scientific) by incubating at room temperature for 15 min.

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

100 µL glycerol stock of the yeast display library was recovered in 50 mL SD-CAA medium by incubating at 27°C with shaking at 250 rpm until OD<sub>600</sub> reached between 1.5 and 2.0. Then 15 mL of the yeast culture was harvested and pelleted via centrifugation at 4000 × g at 4°C for 5 min. The supernatant was discarded, and SGR-CAA (2% w/v galactose, 2% w/v raffinose, 0.1% w/v

D-glucose, 0.67% w/v yeast nitrogen base with ammonium sulfate, 0.5% w/v casamino acids, 0.54% w/v Na<sub>2</sub>HPO<sub>4</sub>, 0.86% w/v NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O, all dissolved in deionized water) was added to make up the volume to 50 mL. The yeast culture was then transferred to a baffled flask and incubated at 18°C with shaking at 250 rpm. Once OD<sub>600</sub> reached between 1.3 and 1.6, 1 mL of yeast culture was harvested and pelleted via centrifugation at 4000 × *g* at 4°C for 5 min. The pellet was subsequently washed with 1 mL of 1× PBS twice. After the final wash, cells were resuspended in 1 mL of 1× PBS.

For expression sort, PE anti-HA.11 (epitope 16B12, BioLegend, Cat. No. 901517) that was buffer-exchanged into 1× PBS was added to the cells at a final concentration of 1 μg/mL. For binding sort of the light chain variant library, PE-conjugated mini-HA was added to washed cells at a final concentration of 30 nM. For Tite-Seq, cells were labeled with PE-conjugated mini-HA at each of five antigen concentrations (one-log increments spanning 0.003 nM–30 nM). A negative control was set up with nothing added to the PBS-resuspended cells. Samples were incubated overnight at 4°C with rotation. Then, the yeast pellet was washed twice in 1× PBS and resuspended in FACS tubes containing 2 mL 1× PBS. Using a BD FACS Aria II cell sorter (BD Biosciences) and FACS Diva software v8.0.1 (BD Biosciences), cells in the selected gates were collected in 1 mL of SD-CAA containing 1× penicillin/streptomycin. Single yeast cells were gated by forward scatter (FSC) and side scatter (SSC). Single cells were then gated by PE anti-HA.11 for expression sort. For Tite-Seq, single cells were gated into three bins along the PE-A axis based on unstained and CR9114 gHC controls, with bin 0 comprising all PE negative cells, bin 2 comprising PE positive cells with comparable expression or binding affinity to the germline CR9114 positive population, and bin 1 comprising the intermediate population between bin 0 and bin 2. Cells were then collected via centrifugation at 3800 × *g* at 20°C for 15 min. The supernatant was discarded. Subsequently, the pellet was resuspended in 100 μL of SD-CAA and plated on SD-CAA plates at 30°C. After 40 h, colonies were collected in 2 mL of SD-CAA. Frozen stocks were made by reconstituting the pellet in 15% v/v glycerol (in SD-CAA medium) and then stored at –80°C until used. FlowJo v10.8 software (BD Life Sciences) was used to analyze FACS data.

### Next-generation sequencing of light chain variant library and CDR H3 library

Plasmids from the yeast cells were extracted using a Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research) following the manufacturer's protocol. The CDR H3 library was subsequently amplified by PCR using primers IGHV1-69-CDRH3-recover-F and IGHV1-69-CDRH3-recover-R whereas the light chain variant library was amplified using primers IGHV1-69-Lightchain-recover-F and IGHV1-69-Lightchain-R. Subsequently, adapters containing sequencing barcodes were appended to the amplicon using primers 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACX XXX XXX XAC ACT CTT TCC CTA CAC GAC GCT-3', and 5'-CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX XXG TGA CTG GAG TTC AGA CGT GTG CT-3'. Positions annotated by an "X" represented the nucleotides for the index sequence. All PCRs were performed using Q5 High-Fidelity DNA polymerase (NEB) according to the manufacturer's instructions. PCR products were purified using PureLink PCR Purification Kit (Thermo Fisher Scientific). The final PCR products were submitted for next generation sequencing using NovaSeq SP PE250 (Illumina).

### Computing the binding score, expression score, and apparent K<sub>D</sub> values

The sequencing data were initially obtained in FASTQ format and subsequently analyzed using a custom python Snakemake pipeline.<sup>31</sup> Briefly, PEAR was used for merging the forward and reverse reads.<sup>32</sup> The number of reads corresponding to each variant in each sample is counted. A pseudocount of 1 was added to the final count to avoid division by zero in downstream analysis. The binding and expression enrichment values of each variant *var* were computed as follows:

$$\text{enrichment}(var) = \log_{10} \frac{\text{Count}_{PE+}(var)}{\text{Count}_{PE-}(var)}$$

where the  $\text{Count}_{PE+}(var)$  is the read count of variant *var* in the PE positive sample for a given binding or expression sort, while  $\text{Count}_{PE-}(var)$  is the read count of variant *var* in the PE negative sample. For the expression sort of the CDR H3 library,  $\text{Count}_{PE+}(var)$  is the read count of variant *var* in bin 2, whereas  $\text{Count}_{PE-}(var)$  is the read count of variant *var* in bin 0. The binding and expression scores for each variant *var* were further computed from the enrichment values as follows:

$$\text{score}(var) = \frac{\text{enrichment}(var) - \overline{\text{enrichment}}_{\text{ve control}}}{\text{enrichment}_{WT} - \overline{\text{enrichment}}_{\text{ve control}}}$$

where  $\overline{\text{enrichment}}_{\text{ve control}}$  is the average enrichment value of the negative controls with stop codons and  $\text{enrichment}_{WT}$  is the enrichment value of the WT. The final score for each variant *var* is the average of two biological replicates.

To compute apparent K<sub>D</sub> value of each CDR H3 variant from the Tite-Seq data, we adopted the analysis approach as previously described.<sup>25</sup> Briefly, to determine the mean bin of PE fluorescence for each CDR H3 variant *var* at each mini-HA concentration, a simple weighted mean calculation was applied:

$$\overline{\text{Bin}}_{[HA]}(var) = \left( \sum_{i=0}^2 N_{i,[HA]}(var) * (i+1) \right) / \left( \sum_{i=0}^2 N_{i,[HA]}(var) \right)$$

where  $N_{i,[HA]}(var)$  is the number of cells with CDR H3 variant *var* that fall into bin *i* at mini-HA concentration [HA]. This calculation computes a weighted average by assigning integer weight to the bin *i*.

For each CDR H3 variant  $var$ , we estimated its sorted cell count  $N_{i,[HA]}(var)$  that corresponds to bin  $i$  at mini-HA concentration  $[HA]$  as follows:

$$N_{i,[HA]}(var) = N_{total,i,[HA]} \times \frac{C_{i,[HA]}(var)}{C_{total,i,[HA]}}$$

where variant read count  $C_{i,[HA]}(var)$  is the read counts for CDR H3 variant  $var$  in bin  $i$  at mini-HA concentration  $[HA]$ ,  $C_{total,i,[HA]}$  is the total read counts for bin  $i$  at mini-HA concentration  $[HA]$ ,  $N_{total,i,[HA]}$  is the total number of cells in bin  $i$  at mini-HA concentration  $[HA]$ .

We then determined the apparent  $K_D$  value for each variant  $K_D(var)$  HA via a nonlinear least-squares regression using a standard non-cooperative Hill equation:

$$\overline{Bin}_{[HA]}(var) = a \times \frac{[HA]}{(K_D(var) + [HA])} + b$$

where free parameters  $a$  is titration response range and  $b$  is titration curve baseline.

CDR H3 variants with a total cell counts of less than 30 across bins 0 to 2 were discarded from our analysis. CDR H3 variants with an apparent  $K_D$  value  $< 10^{-8}$  and a  $p$  value  $> 0.2$  were also discarded. Light chain variants with an average occurrence frequency  $\leq 0.2\%$  across different sorted populations were discarded from our analysis. Sequence logos were generated by Logomaker in Python.<sup>33</sup>

### Expression and purification of Fabs

The heavy and light chains of Fab were cloned into pHCMV3 vector. Light chain mutants were generated using the QuikChange XL Mutagenesis kit (Agilent) following the manufacturer's instructions. The plasmids were co-transfected into Expi293F cells at a 2:1 (HC:LC) mass ratio using ExpiFectamine 293 Reagent (Thermo Fisher Scientific). At 7 days post-transfection, the supernatant was collected, and the Fab was purified using a CaptureSelect CH1-XL Pre-packed Column (Thermo Fisher Scientific).

### Biolayer interferometry binding assay

Binding assays were performed by biolayer interferometry (BLI) using an Octet Red96e instrument (FortéBio) at room temperature as described previously.<sup>34</sup> Briefly, His-tagged mini-HA or SI06 HA proteins at 0.5  $\mu$ M in 1  $\times$  kinetics buffer (1  $\times$  PBS, pH 7.4, 0.01% w/v BSA and 0.002% v/v Tween 20) were loaded onto anti-Penta-HIS (HIS1K) biosensors and incubated with the indicated concentrations of Fabs. The assay consisted of five steps: (1) baseline: 60 s with 1  $\times$  kinetics buffer; (2) loading: 60 s with His-tagged mini-HA proteins; (3) baseline: 60 s with 1  $\times$  kinetics buffer; (4) association: 60 s with Fab samples; and (5) dissociation: 60 s with 1  $\times$  kinetics buffer. Two to three Fab concentrations (33 nM, 100 nM and 300 nM) were used for the assay. For estimating the  $K_D$ , a 1:1 binding model was used. In cases where 1:1 binding model did not fit well due to the contribution of non-specific binding to the response curve, a 2:1 heterogeneous ligand model was used to improve the fitting.

### QUANTIFICATION AND STATISTICAL ANALYSIS

Standard deviation for  $K_D$  estimation was computed by Octet analysis software 9.0. All statistical tests and correlation coefficients were computed in R.