

# Molecular Signatures of Hemagglutinin Stem-Directed Heterosubtypic Human Neutralizing Antibodies against Influenza A Viruses

Yuval Avnir<sup>1</sup>, Aimee S. Tallarico<sup>1</sup>, Quan Zhu<sup>1</sup>, Andrew S. Bennett<sup>1</sup>, Gene Connelly<sup>1</sup>, Jared Sheehan<sup>1</sup>, Jianhua Sui<sup>1</sup>, Amr Fahmy<sup>2</sup>, Chiung-yu Huang<sup>3</sup>✉, Greg Cadwell<sup>4</sup>, Laurie A. Bankston<sup>4</sup>, Andrew T. McGuire<sup>5</sup>, Leonidas Stamatatos<sup>5</sup>, Gerhard Wagner<sup>2</sup>, Robert C. Liddington<sup>4</sup>, Wayne A. Marasco<sup>1\*</sup>

**1** Department of Cancer Immunology & AIDS, Dana-Farber Cancer Institute, Department of Medicine, Harvard Medical School, Boston, Massachusetts, United States of America, **2** Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, United States of America, **3** Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, United States of America, **4** Infectious and Inflammatory Disease Center, Sanford-Burnham Medical Research Institute, La Jolla, California, United States of America, **5** Seattle Biomedical Research Institute, Seattle, Washington, United States of America

## Abstract

Recent studies have shown high usage of the *IGHV1-69* germline immunoglobulin gene for influenza hemagglutinin stem-directed broadly-neutralizing antibodies (HV1-69-sBnAbs). Here we show that a major structural solution for these HV1-69-sBnAbs is achieved through a critical triad comprising two CDR-H2 loop anchor residues (a hydrophobic residue at position 53 (Ile or Met) and Phe54), and CDR-H3-Tyr at positions 98±1; together with distinctive V-segment CDR amino acid substitutions that occur in positions sparse in AID/polymerase-η recognition motifs. A semi-synthetic *IGHV1-69* phage-display library screen designed to investigate AID/polη restrictions resulted in the isolation of HV1-69-sBnAbs that featured a distinctive Ile52Ser mutation in the CDR-H2 loop, a universal CDR-H3 Tyr at position 98 or 99, and required as little as two additional substitutions for heterosubtypic neutralizing activity. The functional importance of the Ile52Ser mutation was confirmed by mutagenesis and by BCR studies. Structural modeling suggests that substitution of a small amino acid at position 52 (or 52a) facilitates the insertion of CDR-H2 Phe54 and CDR-H3-Tyr into adjacent pockets on the stem. These results support the concept that activation and expansion of a defined subset of *IGHV1-69*-encoded B cells to produce potent HV1-69-sBnAbs does not necessarily require a heavily diversified V-segment acquired through recycling/reentry into the germinal center; rather, the incorporation of distinctive amino acid substitutions by Phase 2 long-patch error-prone repair of AID-induced mutations or by random non-AID SHM events may be sufficient. We propose that these routes of B cell maturation should be further investigated and exploited as a pathway for HV1-69-sBnAb elicitation by vaccination.

**Citation:** Avnir Y, Tallarico AS, Zhu Q, Bennett AS, Connelly G, et al. (2014) Molecular Signatures of Hemagglutinin Stem-Directed Heterosubtypic Human Neutralizing Antibodies against Influenza A Viruses. PLoS Pathog 10(5): e1004103. doi:10.1371/journal.ppat.1004103

**Editor:** Peter Palese, Icahn School of Medicine at Mount Sinai, United States of America

**Received** December 7, 2013; **Accepted** March 15, 2014; **Published** May 1, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

**Funding:** This work was supported by the following grants from the National Institutes of Health: NIAID U01-AI074518 to WAM and RCL and NIAID P01 AI094419-01 to LS. This work was also funded by the Defense Advanced Research Projects Agency's "7-Day Biodefense" program under contract # W911NF-10-1-0266 to WAM. The views expressed are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: wayne\_marasco@dfci.harvard.edu

✉ Current address: Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, Maryland, United States of America

## Introduction

Vaccination remains the principle means of preventing seasonal and pandemic influenza and its complications. A “universal” influenza vaccine that induces broad immunity against multiple subtypes of influenza viruses has been a long sought goal in medical research. The discoveries of human broadly neutralizing “heterosubtypic” antibodies binding to various epitopes on the conserved stem domain of HA [1–9] have reignited efforts to develop such a vaccine. These sBnAbs were identified either by panning phage-Ab libraries [1–3,7], or were recovered from B-cells of infected and vaccinated influenza donors [8,10–13]. However, only very low concentrations of sBnAbs are detected in the sera of seasonal influenza [10] or H5N1 vaccinees, or in

commercial intravenous immunoglobulin (**IVIG**) preparations [14]; with notable exceptions being in the response to pdm2009 H1N1 strains and the 1976 swine-flu H1 vaccine [15,16].

sBnAbs are shown to originate from various heavy chain V-segment germline genes. Interestingly, many of these studies (**Table S1 in Text S1**) have reported on the high utilization of the *IGHV1-69* germline gene in broadly neutralizing antibodies against the stem domain of group 1 influenza A viruses (HV1-69 sBnAbs). While *IGHV1-69* germline gene is highly utilized in the population [17], the regulation of this germline gene usage during development and adaptive immune responses has only recently been reexplored [18] following some initial investigations [19–21]. In addition, details of the molecular events that are involved in the elicitation of HV1-69-sBnAbs by vaccination or seasonal influenza

## Author Summary

The quest for universal influenza vaccine has gained wide interest with the discovery of human neutralizing antibodies that are able to variably cross neutralize and protect against different influenza strains, subtypes, groups and lineages. These antibodies, which bind to a highly conserved epitope in the hemagglutinin stem, are often encoded by rearranged *IGHV1-69* germline genes that alone make contact with HA and prevent virus entry and emergence of escape mutants. Our study was undertaken to gain an understanding of what structural requirements enable a rearranged *IGHV1-69* Ab to become a potent cross-neutralizing antibody. We found that in addition to a critical amino acid triad consisting of a pair of anchor residues in CDR-H2 and a properly positioned CDR-H3 Tyr, distinctive V-segment substitutions that arise in positions that are distinct from phase I AID somatic hypermutation (SHM) hotspot motifs are often required. As few as two V-segment SHM can fulfill this role which appears to facilitate the optimal binding of CDR-H2 Phe54 and CDR-H3-Tyr into adjacent hydrophobic pockets in the HA stem. These studies provide new information on the SHM requirements for *IGHV1-69*-encoded B cells to produce HV1-69-sBnAbs and suggest that there may exist alternative routes to their elicitation by vaccination.

infection remains unknown. The highly immunogenic globular head [1,10,11] is thought to be a main impediment for sBnAb elicitation as the stem epitopes have been shown to be readily accessible to sBnAbs [22].

In this study we sought to better define the V-segment amino acid substitutions and CDR-H3 amino acids within rearranged *IGHV1-69* germline genes that are preferentially used to allow an *IGHV1-69* germline based Ab to become a potent HV1-69-sBnAb. Analysis of 38 HV1-69-sBnAbs recovered from 8 laboratories (**Table S1 in Text S1**) indicates that broad-spectrum binding and neutralization is conveyed by a triad of critical anchor residues composed of two CDR-H2 residues including a hydrophobic residue at position 53 and Phe54, and properly positioned CDR-H3 tyrosines. In addition, we define distinctive V-segment mutations within the CDR H1/H2/H4 loops. Moreover, these V-segment mutations occur in positions that are sparse in activation-induced cytidine deaminase (AID) and polymerase eta (pol η) consensus “hot-spot” motifs. Together with panning of a semi-synthetic *IGHV1-69* Ab library against H5/H1 HAs, mutagenesis studies, and structural modeling we demonstrate that HV1-69-sBnAbs can be evolved from a *IGHV1-69* germline gene with as few as two V-segment substitutions with one occurring at CDR-H2 Ile52Ser, and by properly positioned Tyr in the CDR-H3 domain. The CDR-H2 substitutions at positions Ile52Ser and Pro52aGly/Ala are predicted to function not by making new contacts with the epitope themselves, but rather by enabling conformational changes within the CDR loops that facilitate optimal insertion of two major anchor residues CDR-H2 Phe54 and CDR-H3-Tyr98 into adjacent pockets in the stem. Our immunogenetic and structural studies demonstrate that the generation of critical SHM for HV1-69-sBnAbs does not occur through the classical phase I AID repair mechanism that takes place directly under WRCY/RGYW motifs, instead by phase 2 long-patch error-prone repair or random non-AID SHM events. Further, these results suggest that the secondary AID repair mechanisms as described here may not require B cell recycling/reentry in the germinal center [23], rather by alternative routes such as short term entry into the germinal center or in a

specialized extra-follicular location such as the marginal zone [24].

## Results

HV1-69-sBnAb CDR-H3 Tyrs together with CDR-H2 Phe54 mediate high affinity binding to adjacent pockets in the HA stem

The co-crystal structures of the HV1-69-sBnAbs, F10 [1], CR6261 [4], and CR9114 [7] with H5VN04 established that binding is mediated exclusively by the *IGHV1-69* heavy chains. To further explore binding similarities in the co-crystal structures, binding free energy contributions for the heavy chain CDR residues were estimated by using the ANCHOR server [25] (**Figure 1A**). This analysis revealed the under recognized importance of CDR-H3-Tyr98 as another common anchor residue to the already two known hydrophobic residues at position 53 (Ile/Met), and the conserved Phe at position 54 [1,4,26]. In these three HV1-69-sBnAb structures Tyr98 is located in close proximity to Phe54 (~4 Å) and is the only contact residue in which the side chain adopts a single conformation in its binding pocket (**Figure 1B**), and forms a strong H-bond with the fusion peptide (the main chain carbonyl of Asp19<sub>2</sub>). To further validate the importance of CDR-H3 Tyr98, the binding kinetics of F10 and CR6261 Y98A variants to H5VN04 were compared to wt and F54A variants (**Figure 1C**). Binding kinetic studies showed that Y98A reduced binding by >3000-fold and >4,000-fold for F10 and CR6261, respectively. In addition, the F54A substitution completely ablated binding of F10 and reduced the CR6261 binding by >800-fold, which is consistent with previous studies [3,26]. Thus, the poor binding kinetics of the Y98A variants further substantiates CDR-H3 Y98 as a major contributor to HA binding.

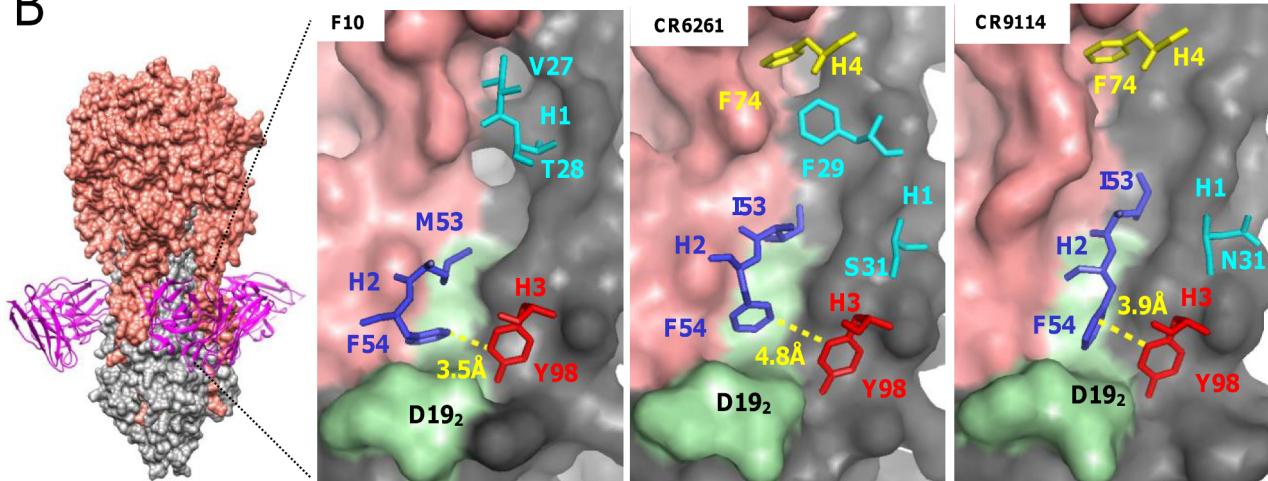
### Identifying distinctive CDR-H3 and V-segment molecular signatures in HV1-69-sBnAbs

The conserved triad of a hydrophobic residue at position 53, Phe54, and Tyr98, led us to explore the commonality of these residues in published HV1-69 sBnAbs (**Table S1 in Text S1**). **Figure 2A** shows that hydrophobic residues are always found at CDR-H2 position 53, Phe54 is nearly invariant, and CDR-H3 tyrosines are found in 35/38 HV1-69 sBnAbs. Further immunogenetic analysis demonstrated that 37 of these HV1-69 sBnAbs belong to the *IGHV1-69* 51p1 allele group, all of which encode the critical Phe54 (**Figure S1A-B in Text S1**) whereas the D-segments or J-segments are highly diverse (**Figure S1C-D in Text S1**).

To identify possible distinctive V-segment amino acid substitutions in the HV1-69-sBnAb Ab dataset, we compared all of their amino acid substitutions with a reference V-segment dataset composed of functional, non-duplicated *IGHV1-69*-51p1 allele based Abs (IgBlast, n = 287). Thirteen distinctive HV1-69-sBnAb V-segment substitutions were identified using Fisher's exact test with Bonferroni adjustment (highlighted in **Figure 2A** and respective frequencies shown in **Figure 2B**). When the HV1-69-sBnAbs are ordered according to the occurrence of the distinctive CDR-H2 substitutions, a prominent cluster of Ser52, Gly52a, and Ala52a mutations is seen in 21/27 antibodies with CDR-H3 domains that are also characterized by tyrosines in positions 97-to-99 (**Figure 2A**). The Gly52a substitutions are associated with dual or triple CDR-H3, and Ala52a substitutions are strongly associated with the distinctive substitution of Arg30 (6/8). Furthermore, distinctive CDR-H4 (**Figure S2 in Text S1**) substitutions of Gln73 and Phe74, and CDR-H1 Pro28 and Arg30 only occur with HV1-69-sBnAbs containing CDR-H3-Tyr97-99.

**A**

	CDR-H1	33	51	CDR-H2	54	57	73	76	92	CDR-H4	98	CDR-H3	101
IGHV1-69*01	GGT FSS YA - I I P I FG TA - E S TS - C AR												
F10	E V T F S S FA - I S P M F G T P - Q S TR - C A R S P S Y I C S G G T C V F D H W												
CR9114	GG T S N N YA - I S P I F G S T - I F S N - C A R H G N Y Y Y S G M - - - D V W												
CR6261	GG P F R S YA - I I P I F G T T - D F A G - C A K H M G Y Q V R E T M - - - D V W												

**B****C**

Variant	H5VN04 binding			
	ka H5*	kd H5*	KD H5*	KD variant/ KD wt
F10 wt	$3.4 \times 10^{-5}$	$3.5 \times 10^{-5}$	$1.0 \times 10^{-10}$	1
F10 F54A	Non-detectable			
F10 Y98A	$2.8 \times 10^{-5}$	$9.7 \times 10^{-2}$	$3.5 \times 10^{-7}$	3305
CR6261 wt	$3.5 \times 10^{-5}$	$1.2 \times 10^{-4}$	$3.4 \times 10^{-10}$	1
CR6261 F54A	$2.5 \times 10^{-14}$	$7.3 \times 10^{-3}$	$2.9 \times 10^{-7}$	885
CR6261 Y98A	$1.7 \times 10^{-14}$	$2.7 \times 10^{-2}$	$1.6 \times 10^{-6}$	4839

\*- Kinetic values are rounded to the nearest tenth.

**Figure 1. The structural basis of HV1-69-sBnAb similarity.** **A)** The ANCHOR web server [25] was used to identify heavy chain CDR residues that make favorable contacts ( $-1$  kcal/mol  $>-3$  kcal/mol orange) and highly favorable binding contacts ( $<-3$  kcal/mol red) in the co-crystal structures of F10 (PDB: 3FKU) CR6261 (PDB: 3GBM), and CR9114 (PDB: 4FQI). **B)** Left - The location of F10 binding on the HA is shown with HA1 colored in salmon and HA2 colored in grey. Right panels – The location of the CDR residues identified in **A)**. In light green is the HA2 fusion peptide from Trp21<sub>2</sub>-to-Val18<sub>2</sub>. **C)** Binding kinetics data of F10, CR6261, and the respective variants of F10 F54A, F10 Y98A, CR6261 F54A and CR6261 Y98A, against H5VN04.

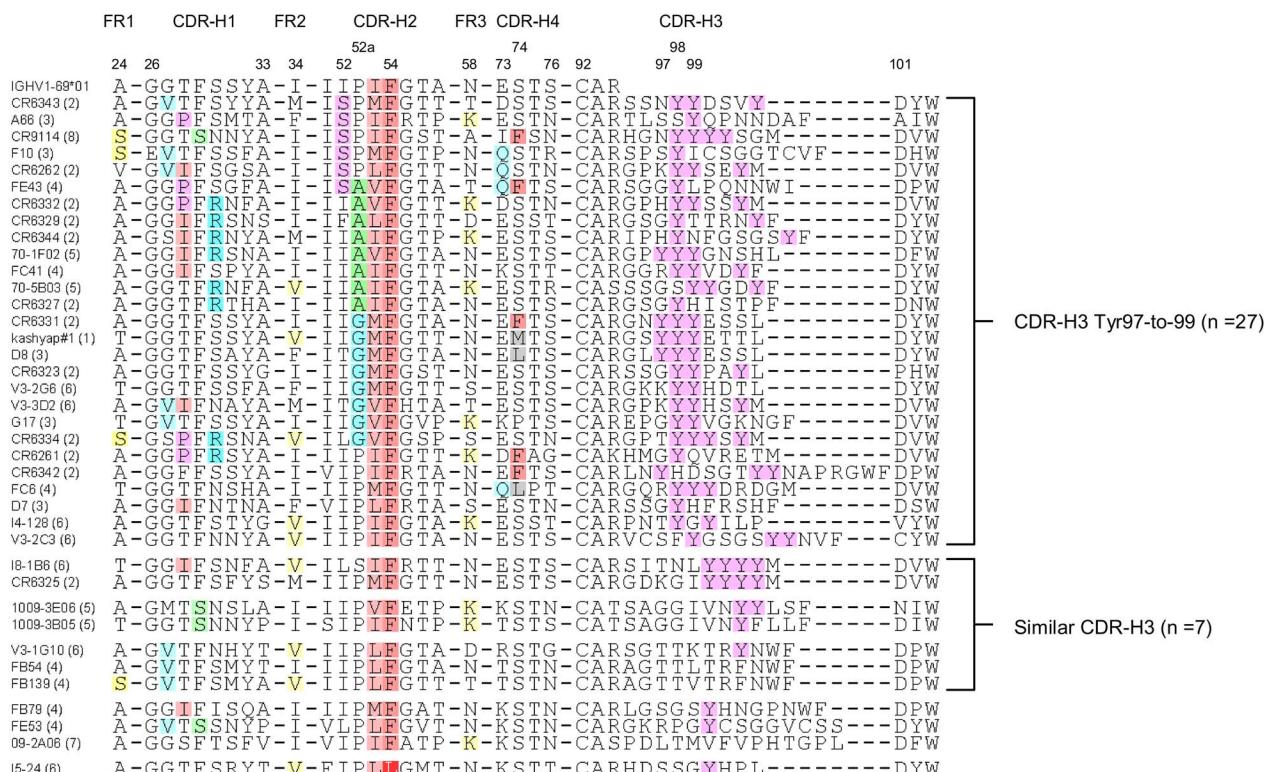
doi:10.1371/journal.ppat.1004103.g001

Of the other 10 HV1-69-sBnAbs not characterized by CDR-H3-Tyr97-to-99, several high CDR-H3 sequence similarities were observed, which allowed grouping of 7 of these HV1-69-sBnAbs into three sub-clusters.

### Studying the origins of the distinctive HV1-69-sBnAb V-segment CDR substitutions

**Figure 3A** shows that HV1-69-sBnAbs are characterized by a mean of  $12.6 \pm 4.2$  V-segment substitutions that ranges in a continuum from 5 in CR6331/CR6432 to 22 in FE43/CR6334. This is in marked distinction to the much higher number of somatic mutations that are found in anti-HIV gp120 CD4-binding-site sBnAbs that show biased use of the *IGHV1-02* germline gene [27]. Furthermore, the distinctive substitutions are distributed across this continuum as exemplified by the CDR-H2 Ser52 and Gly/Ala52a substitutions.

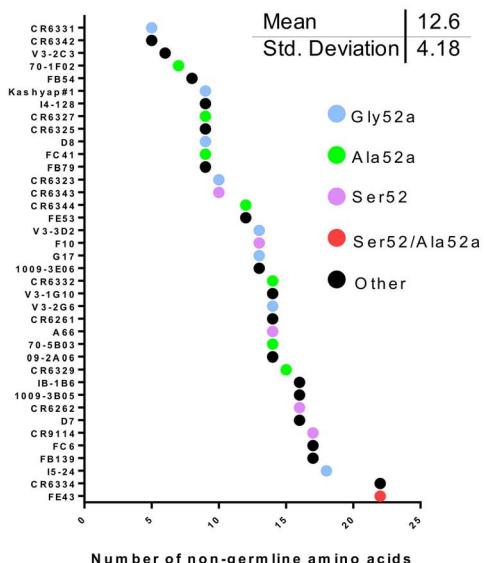
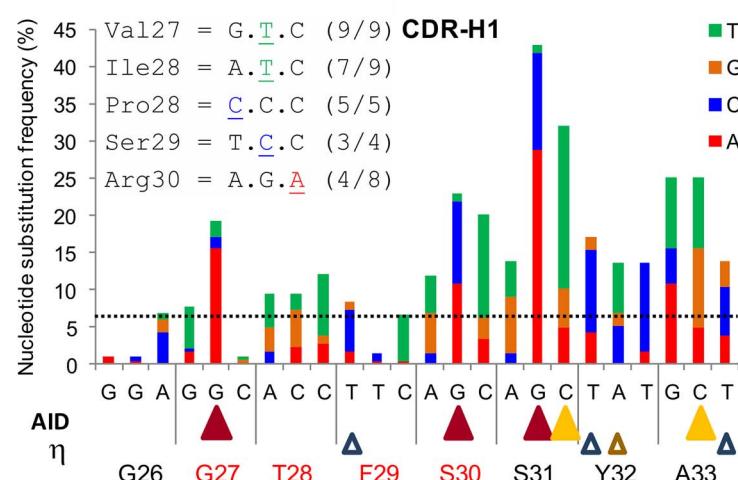
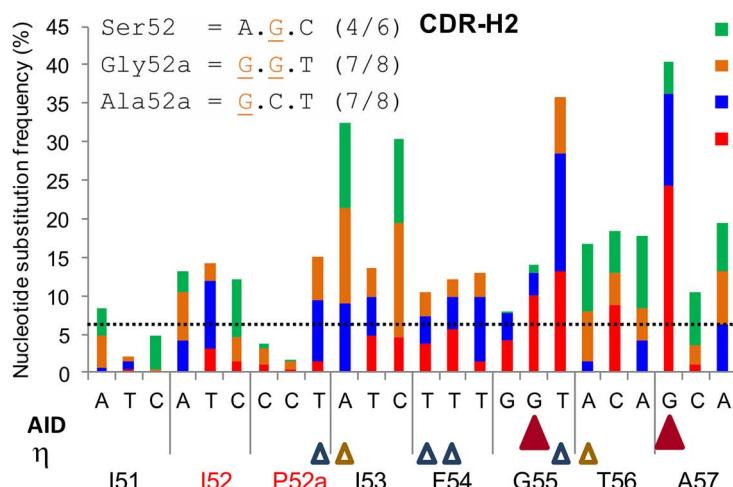
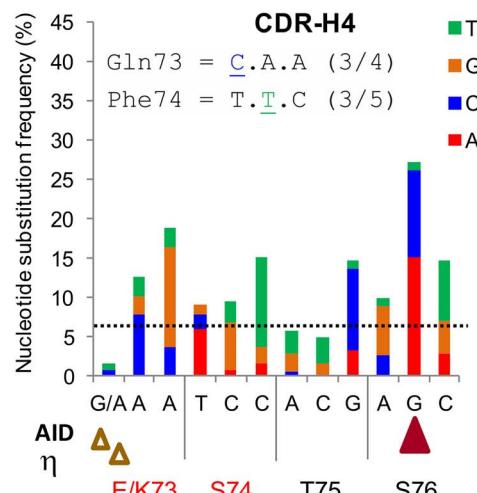
The nucleotide changes in CDR H1, H2 and H4 that are responsible for these distinct HV1-69 sBnAb substitutions are shown to occur mostly by one nucleotide and not multiple nucleotide changes in the codon (**Figure 3B-D upper panel inserts**). This observation led us to further explore the nucleotide substitution frequencies in the same non-duplicated *IGHV1-69* reference dataset and its relation to the location of AID and pol η “hotspot” somatic hypermutation (SHM) motifs [28,29]. In agreement with the study by Clark et al [30], positions of high nucleotide substitution frequency are mostly found directly under AID hotspot motifs. It is also noticeable that AID motifs are found to concentrate towards the 3' end of the CDR loops. In CDR-H1 4/5 AID hotspots are located 3' to position 29, in CDR-H2 2/2 are located 3' to position 54 and in CDR-H4 1/1 is located 3' to position 75. Notably, of 10 distinctive substitutions on CDR-H1,H2 and H4 only two, Val27 and Arg30 are directly under AID

**A****B**

AA substitution	IGHV1-69-Ab dataset (n = 287) (%)	HV1-69-sBnAb dataset (n = 37) (%)	P <sup>a</sup>
<b>FR1</b>			
A24S	0.4%	11%	0.0047
<b>CDR-H1</b>			
G27V	0.7%	24%	<0.0001
T28I	1.7%	24%	<0.0001
T28P	1.1%	14%	0.0065
F29S	0.4%	11%	0.004
S30R	4.2%	22%	0.0057
<b>FR2</b>			
I34V	6.3%	22%	0.0185
<b>CDR-H2</b>			
I52S	0.4%	16%	0.0001
P52aG	0.7%	22%	<0.0001
P52aA	1.1%	22%	<0.0001
<b>FR3</b>			
N58K	7.7%	27%	0.0153
<b>CDR-H4</b>			
E73Q	0.4%	11%	0.0067
S74F	1.7%	14%	0.0222

<sup>a</sup> - P-values were obtained by Fisher's exact test with Bonferroni adjustment.

**Figure 2. Characterization of HV1-69-sBnAbs VH domain. A)** Alignment of 38 published HV1-69-sBnAbs is shown with highlights referring to hydrophobic residues at position 53 (light plum), the conserved Phe54 (dark plum), the occurrence of CDR-H3-Tyr (pink) residues. Other highlights refer to panel B, which describes the result of a Fisher's exact test with Bonferroni adjustment that compared V-segment amino acid substitutions diversity and frequency of the 37 51p1 allele related HV1-69-sBnAbs with that of a reference IGHV1-69 51p1 allele related Ab dataset. 13 amino acid substitutions were determined to uniquely associate with the HV1-69-sBnAb dataset ( $P < 0.05$ ). doi:10.1371/journal.ppat.1004103.g002

**A****B****C****D**

**Figure 3. Studying the location of SHM hotspots and nucleotide substitutions frequencies in the *IGHV1-69* reference Ab dataset. A)** Number of V-segment substitutions observed in 38 HV1-69-sBnAbs with color notations that designate the HV1-69-sBnAbs characterized by the distinctive CDR-H2 Ser52 or Ala/Gly52a and “other” which do not contain CDR-H2 Ser52 or Ala/Gly52a. **(B-D) Upper panel – the common reference dataset was studied for substitution frequency of nucleotides in the V-segment CDRs and for location of AID and polη hotspots (AID=WRCY yellow solid triangle/RGYW dark red solid triangle, Polη WA brown open triangle/TW blue empty triangle. The horizontal line shows the mean $\pm$ SD (6.44 $\pm$ 7.23) of non-germline nucleotide substitution frequency observed for FR regions to serve as a reference.**

doi:10.1371/journal.ppat.1004103.g003

motifs. The Gly27Val substitution has the unusual property of occurring by dG-to-dT transversion instead of the phase Ia AID associated dG-to-dA transition to Gly27Asp, suggesting it occurred by a phase Ib short patch base excision repair (BER) mechanism and/or by long patch (BER)/mismatch repair (MMR) phase 2 mechanism [28]. The polη (open triangles) SHM motifs are more

disperse, however the 5' end of both CDR-H1/H2 domains is also sparse of this SHM motif. Hence, it appears that in response to influenza infection or vaccination mutations in the V region are not the result of the direct biochemical action of AID, but rather depend on the error-prone BER and MMR of the AID-induced mutation.

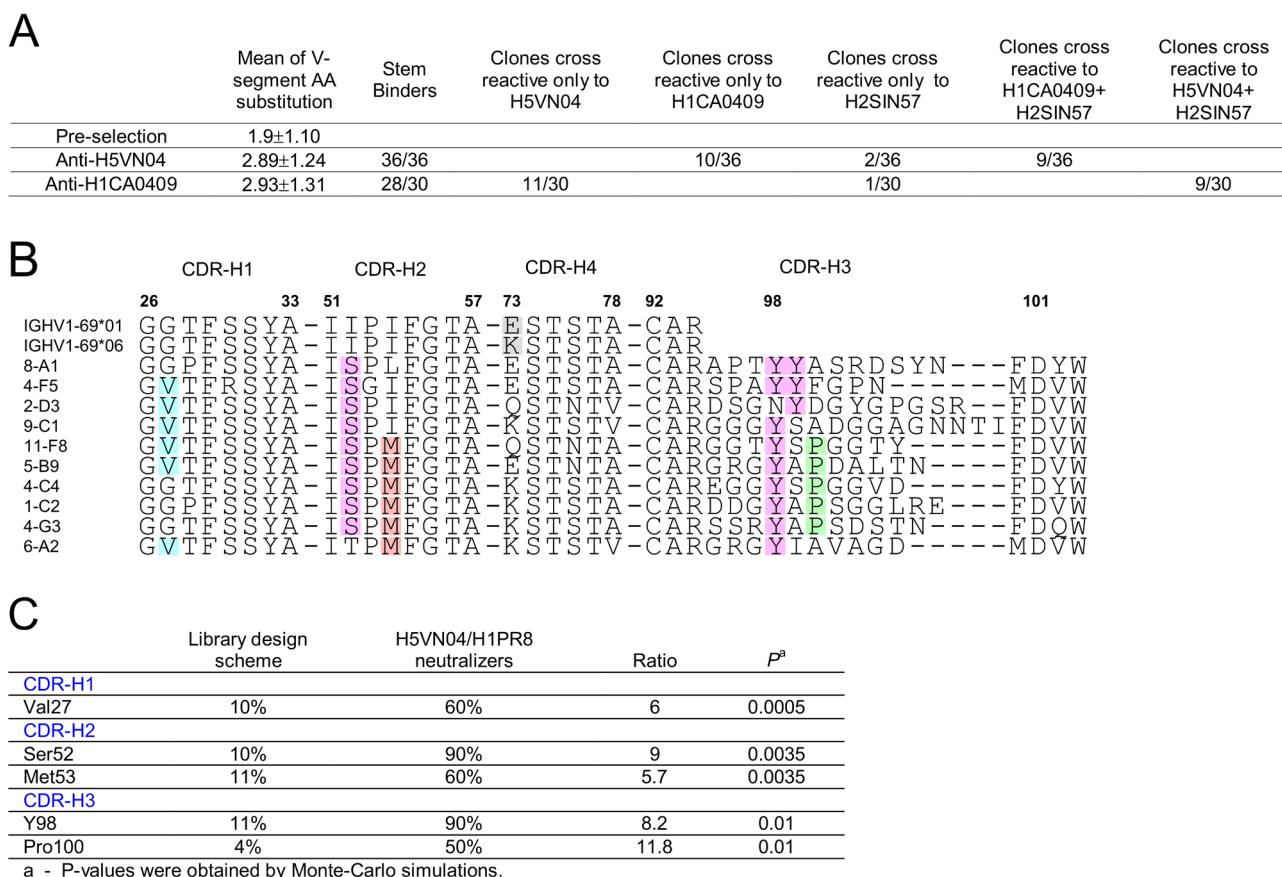
## Exploring the importance of distinctive HV1-69-sBnAbs substitutions through use of a semi-synthetic *IGHV1-69* library that bypasses AID/Pol η restrictions

In consideration of the immunogenetic DNA repair mechanisms that may be responsible for the critical CDR-H1/H2/H4 positions in HV1-69 sBnAbs noted above, we generated a semi-synthetic phage-antibody library (5 displayed Abs/phage) to bypass restrictions that may be related to the SHM machinery. The library was designed with a low V-segment amino acid substitution frequency ( $1.9 \pm 1.1$ ) and incorporated 9 of the 13 distinctive HV1-69 sBnAb amino acid substitutions at a frequency no higher than 10% and with a completely randomized CDR-H3 of varying length (**Supporting text in Text S1**). The library is strongly skewed towards selection of Ab-members that display germline residues. For example, the combination of the V-segment germline residues of CDR-H1 Gly27 (90%) and CDR-H2 Ile52 (71%) with CDR-H3 Tyr98 (11%) is expected to occur in 7% of the phage members whereas the combination of the distinctive HV1-69-sBnAb substitutions of Val27 (10%), Ser52 (10%) with CDR-H3 Tyr98 is (11%) expected to occur in 0.11% of the phage members (**Figure S3 in Text S1**).

Panning the library against the H5-VN04 or H1CA0409 trimeric HA proteins resulted in the isolation of 36 and 30 unique phage-Ab clones, characterized by low V-segment amino acid

substitution frequency of  $2.89 \pm 1.24$  and  $2.93 \pm 1.31$ , respectively (**Figure 4A**). F10 competition assays performed with H1CA0409, H2SIN57 or H5VN04 coated on MSD plates further indicated that all anti-H5VN04 phage Ab are stem directed and 10 cross-reacted with H1CA0409; 2 cross-reacted with H2 A/Singapore/1/1957 (H2SIN57); and 9 cross-reacted with both H1CA0409 and H2SIN57 (**Figure S4A in Text S1**). Likewise, 28/30 anti-H1CA0409 phage Abs target the stem, and 11 cross-reacted only with H5VN04; 1 cross-reacted only with H2SIN57; and 9 cross-reacted against both H1CA0409 and H2SIN57 (**Figure S4D in Text S1**). Of the two non-stem binder only one cross-reacted against H5VN04.

Heterosubtypic neutralization activity was tested for thirty-one anti-H5VN04-stem phage-Abs by using H5VN04 and H1PR8 pseudotyped viruses. Ten of these phage-Ab characterized by >95% neutralization activity against both strains are shown in **Figure 4B**. Chi<sup>2</sup> based statistical analysis of these 10 sBnAbs that compares the frequency of substituted amino acids before and after selection (**Figure 4C**) revealed a sequence solution conveyed by two distinctive HV1-69-sBnAb substitutions CDR-H1 Val27 (6/10) and CDR-H2 Ser52 (9/10) together with CDR-H3 CDR-H3-Tyr 98 (9/10). Overlapping non-HV1-69-sBnAb specific substitutions of Met53 (6/10) and CDR-H3 Pro100 (5/10) were also found in this pool. In the remaining pool of anti-H5VN04



**Figure 4. Semi-synthetic HV1-69 phage-Ab library yields potent anti-H5VN04/H1CA0409 Abs characterized by a minimal V-segment amino acid substitutions.** A) Characterization of binding activities of anti-H5VN04 and anti-H1CA0409 phage-Abs isolated from the semi-synthetic HV1-69 phage-display library. Sequences are detailed in **Figure S4 in Text S1**. B) Heavy chain CDR sequences of anti-H5VN04 phage Abs characterized by >95% neutralization activity against both H5VN04 and H1PR8 pseudotyped viruses. The 5 highlighted residues in the CDRs refer to panel C) which describes the result of a Chi square statistical analysis approach used to identify residues that were significantly enriched as compared to their frequency in the library ( $P < 0.05$ ). Also highlighted are Tyr99 and position73 in the *IGHV1-69* germline sequences.  
doi:10.1371/journal.ppat.1004103.g004

phage-Abs (**Figure S4B in Text S1**), the substitutions of CDR-H4 Asn76 and CDR-H3 Gly97, Tyr99, and Gly100B were also significantly enriched.

**Figure S4A in Text S1** also shows that the dominant sequence motif of Ser52/Tyr98 occurs in both the heterosubtypic (16/21) and the non-heterosubtypic anti-H5VN04 phage-Ab subsets (8/15). In order to understand if heterosubtypic activity of phage-Abs characterized by the Ser52/Tyr98 sequence motif is associated with other amino acids, the composition of the CDR-H3 domain of the two subsets was analyzed separately for the occurrence of enriched residues. The statistical analysis in **Figure S4C in Text S1** shows that in the heterosubtypic Ser52/Tyr98 subset Tyr99, Pro100, and Gly100B were significantly enriched whereas no significant enrichment of these same residues were found to occur in the CDR-H3 of the non-heterosubtypic subset. Furthermore, although not statistically significant, the high frequency of glycines in the heterosubtypic subset is also shown in **Figure S4A in Text S1** to occur 5' to CDR-H3-Tyr98 where 46% of the amino acids at positions 95-to-97 are glycines as opposed to the non-heterosubtypic subset where 21% are glycines at these positions. This analysis suggests that a flexible CDR-H3 loop is beneficial in mediating heterosubtypic activity for anti-H5VN04 stem binders characterized by CDR-H2 Ser52 and CDR-H3-Tyr98.

A similar amino acid enrichment profile was also observed in the H1CA0490 phage-Ab pool (**Figure S4D–E in Text S1**) that are characterized by the dominant pair of Ser52/Tyr98 as well as by CDR-H1 Val27, CDR-H4 Asn76 and CDR-H3 Gly97. These substitutions are also shown to dominate the heterosubtypic phage-Ab subset whereas in the non-heterosubtypic subset CDR-H1 Val27 and CDR-H2 Ser52 appear only once (1/9) and Tyr residues in positions 98 or 99 appear only in three phage-Abs (3/9).

The predominant recovery of Ser52 over Gly52a and Ala52a encoding phage-Abs from the panning campaigns despite similar coding frequency in the library (**Figure S3 in Text S1**) was unexpected but might be explained in view of the fact that the Gly52a subset is restricted to double or triple tyrosines in the CDR-H3 domain, whereas the Ala52a subset is shown to be strongly associated with CDR-H1 Arg30 (6/8) (**Figure 2A**). In contrast, the Ser52 does not appear to be as strongly associated with other V-segment substitutions. Thus, the incorporation of Ser52 likely provides a higher diversity of structural solutions than that of Gly52a and Ala52a as follows: the frequency of phage-Ab members characterized by CDR-H2Ser52/CDR-H3-Tyr98 is expected to be 1%, the frequency of phage members characterized by CDR-H2 Gly52a/CDR-H3-Tyr98 and 99 is expected to be ~0.14% and the frequency of phage members characterized by CDR-H1 Arg30/CDR-H2 Ala52a and CDR-H3-Tyr98 is ~0.11%.

### Confirming the importance of the CDR-H2 substitutions in HV1-69-sBnAbs

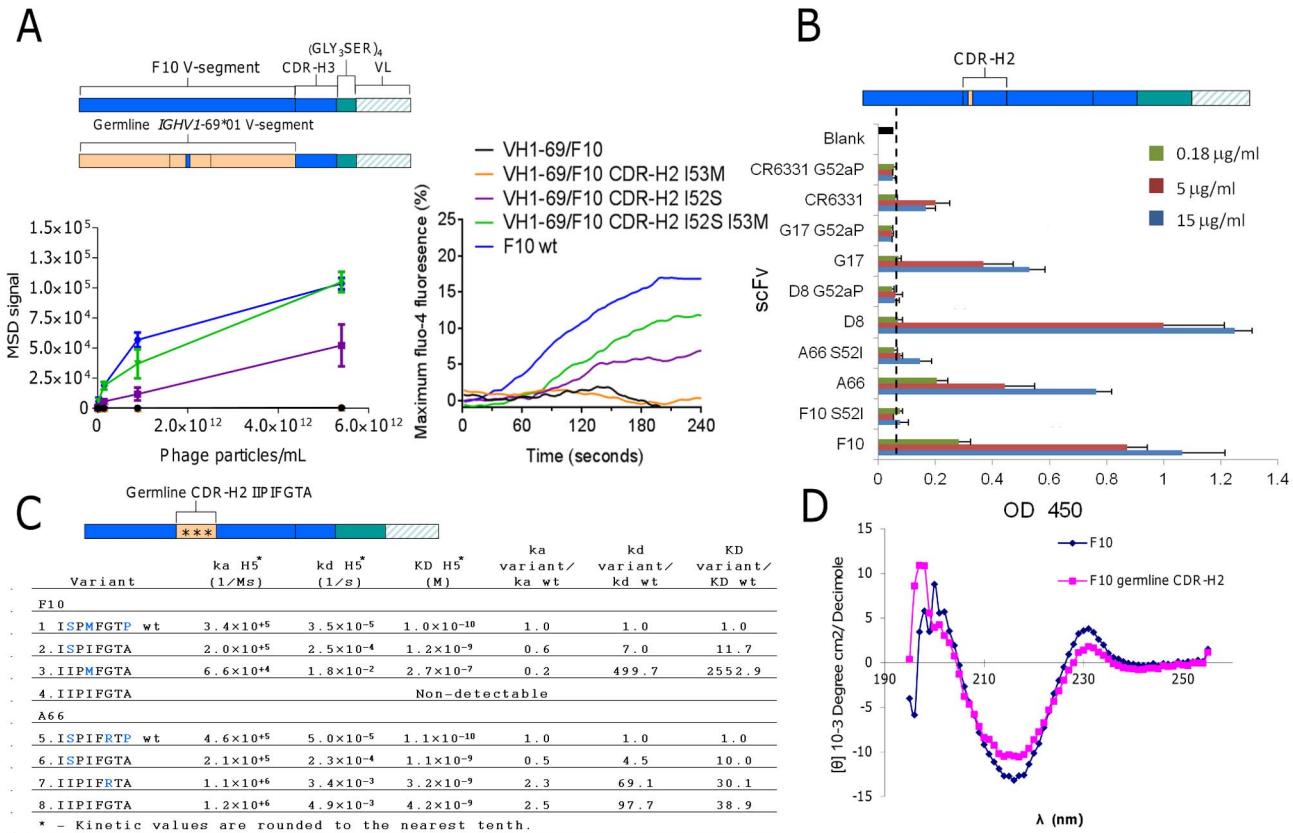
Since Ser52 was the dominant sequence solution obtained from the panning campaigns of the synthetic library we next sought to test the relative importance of Ser52 in F10 when the V-segment was converted to the non-mutated *IGHV1-69\*01* form (VH1-69/F10) and then back introduced with Ser52 and Met53 separately and together. In order to utilize avidity to increase detection of weak interactions, the VH1-69/F10 variants were either expressed on the surface of phage particles and binding tested with an MSD ELISA assay (**Figure 5A, left**); or expressed as B-cell receptors and analyzed for their ability to activate B-cells through cross-linking with purified H5VN04 trimeric HA (**Figure 5A, right**). In both cases, we detected no binding of H5VN04 with either the germline or the I53M variant. However, the I52S variant was active, and the I52S/I53M variant had even higher activity.

To further validate the structural role of the distinctive CDR-H2 substitutions in the context of the fully mutated HV1-69-sBnAbs we mutated five representative HV1-69-sBnAbs carrying either I52S or P52aG mutation back to the germline gene residue. All five variants had drastically reduced or abolished binding reactivity to H5VN04 (**Figure 5B**). In comparison, a revertant proline substitution at CDR-H2 position 57 in F10 and A66 resulted in enhanced or no change in H5 binding, respectively (data not shown). We also constructed F10 and A66 CDR-H2 germline variants then back introduced the substitutions of Ser52 (F10/A66), Met53 in F10, and Arg55 in A66 (**Figure 5C**). The kinetic data shows only small differences in association rate ( $k_a$ ) constants among the wild-type and variants, whereas much greater effects were observed with the dissociation rate ( $k_d$ ) constants. Replacing wild-type CDR-H2 with the germline sequence (IPIIFGTA) led to a 98-fold higher dissociation rate in A66 (construct #5 vs. #8) and no detectable binding activity in F10 (construct #1 vs. #4). Circular dichroism (CD) analysis indicated that protein misfolding is not responsible for loss of binding as a similar profile was observed for the F10 germline CDR-H2 variant and F10 wt (**Figure 5D**). Restoring the single amino acid of Ser52 (ISPIIFGTA) (mutants #2 & 6) resulted in recovery of binding kinetics for both sBnAbs, as seen by the dramatic improvement in  $k_d$  values. Other single amino acid CDR-H2 reversions (F10 IIPMFGT, A66 IIPIFRTA) (mutants #3 and 7) did not restore binding to the same extent and showed extremely fast  $k_d$  rates (**Figure 5C**).

### Understanding the structural role of the HV1-69-sBnAb distinctive CDR-H2 substitutions

Ser52 of F10 and CR9114 do not form high energy contacts with the respective H5VN04 HAs, as evidenced by van der Waals (VDW) contact analysis (**Figure 6A**); thus, direct contacts with antigen are not the source of the dramatic effect of the Ile52Ser mutation. Analyzing the contacts between Phe54 and Tyr98 in the H5VN04-bound structures shows Phe54 making close (~3.7 Å) orthogonal contacts with Tyr98 in the F10 and CR9114 complexes, whereas for CR6261, the 2 rings are further apart (4.8 Å) and nearly coplanar. *In silico* mutagenesis of Ser52 to germline Ile52 in a F10 model in which its amino acid side chains are allowed to rotate, shows an increase in number of VDW contacts (**Figure S5 in Text S1**). Furthermore, it was noted that Ile52 could not be accommodated in the F10 and CR9114 H5-bound structures in which the amino acids side chains were fixed. This can be visualized by structural alignment of the CDR-H2 loops of F10, CR9114 and CR6261, which shows Ile52 sterically clashing with both F10 and CR9114 (**Figure 6A, right panel**). In addition, structural alignment between unbound CR9114 and the H5-bound structure (**Figure 6B**) shows that only CDR-H2 residues, Ile53 and Phe54, adopt markedly different positions (>1.7 Å shifts) following HA binding. Distance analysis suggests an induced-fit process, in which the  $\text{C}\alpha\text{-C}\beta$  atoms of Phe54 are shifted closer to the CDR-H3 Tyr98 domains in the bound state. We propose that this induced fit process could not occur if the large germline Ile52 is maintained in that position.

The close distance between Phe54 and CDR-H3-Tyr is also expected to be a structural feature in HV1-69-sBnAbs carrying the P52aG/A substitutions, since, like I52S, they are characterized by a narrow distribution of CDR-H3 Tyr residues clustered around positions 97-to-99 (**Figure 2A**). *In silico* mutagenesis suggests that replacing germline Pro52a in the non-mutated *IGHV1-69* germline-based Ab 1-69/B3 [31] with Gly52a also leads to a reduction in VDW contacts (**Figure S5 in Text S1**).



**Figure 5. Validating the structural role of Ser52 in HV1-69-sBnAbs.** **A**) F10 V-segment germline variants were analyzed for H5N04 binding in the phage-Ab (5 scFv/phage) format by MSD ELISA (left) and for their ability to activate B-cell when expressed as B-cell receptors in the presence of H5N04 (right). **B**) HV1-69-sBnAb variants of S52I in F10 and A66, G52aP in CR6331, G17 and D8 were analyzed for H5N04 reactivity by ELISA. **C**) Kinetic analysis by Biacore of F10 and A66 CDR-H2 variants against purified H5N04. Residues colored in blue are non-germline amino acids. **D**) Circular dichroism measurement of F10 and the non-H5 reactive variant characterized by a germline configured CDR-H2 shows a highly similar CD profile for both constructs.

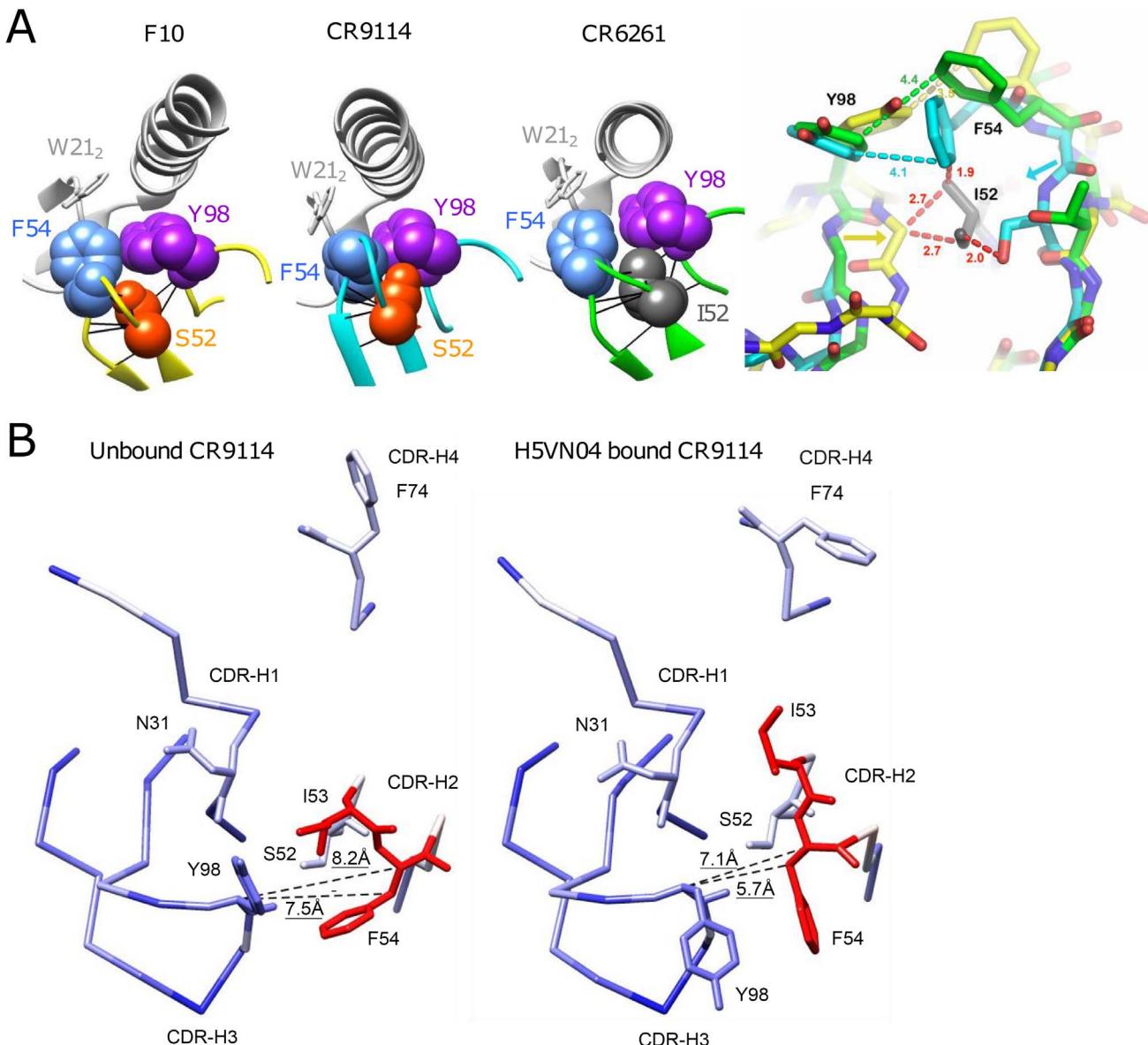
doi:10.1371/journal.ppat.1004103.g005

## Discussion

Development of a “universal” influenza vaccine that is aimed at the elicitation sBnAbs should remain a high priority for global health and national biosecurity. Such a vaccine should preferably induce sBnAbs of high diversity that target various stem epitopes and would not allow generation of escape variants. In this respect the *IGHV1-69* germline gene based sBnAbs have the property of showing recalcitrance to neutralization escape [1,3,10]. This feature and the isolation of HV1-69-sBnAbs from various donors led us to further characterize reported HV1-69-sBnAbs and understand what are the requirements for *IGHV1-69* germline based Abs to become a potent HV1-69-sBnAb. Starting with a comparative structural analysis of three published HV1-69-sBnAbs, our analysis shows that a shared CDR-H3-Tyr98 assumes identical conformations within its own stem pocket and forms a strong H-bond with the fusion peptide (**Figure 1A–B**). Functional importance of Tyr98 was shown by mutagenesis studies and by their high frequency of recovery from panning studies and is also implied from the occurrence of the CDR-H3-Tyr 97–99 cluster in 71% of the reported HV1-69-sBnAbs (**Figure 2A**). Hence, this analysis adds a third common anchor residue to the two well described contact amino acids of hydrophobic Met/Ile53 and Phe54, and provides a new view on the contribution of the CDR-H3 loop on HA binding [4,32].

Immunogenetic analysis shows that HV1-69-sBnAbs CDR-H3 Tyr98 arise from a diverse repertoire of D-segments, whereas 37/38 of the HV1-69-sBnAbs are shown to utilize *IGHV1-69* V-segment alleles (6/13) characterized by CDR-H2 Phe54 (i.e., the 51p1 allele group). This allelic bias cannot be attributed to the frequency of 51p1 and hv1263 alleles in the population [33] and further supports a critical role of Phe54 (**Figure S1 in Text S1**). These results suggest that precursor pool of HV1-69-sBnAbs preferentially requires the triad of anchor residues of CDR-H2 Ile53/Met53, CDR-H2 Phe54 and CDR-H3-Tyr98. Based on these immunogenetic findings, it appears that HV1-69-sBnAb elicitation may be constrained when an individual is 51p1-null and expresses only Leu54 from the hv1263 alleles (circa 20% of the population) [33] (Avnir et al in preparation). We observed a low frequency (circa 6–7%) of Leu54Phe substitutions in the hv1263 allelic group by NGS sequencing of our 27 billion member non-immune Ab-phage library (data not shown).

Sequence similarities among the HV1-69-sBnAbs were also observed in the V-segments. Our statistical analysis identified thirteen amino acid substitutions that are distinctive to the HV1-69-sBnAb dataset when compared to *IGHV1-69* 51p1 germline based Abs obtained from the IgBlast database. Seven of these changes: CDR-H1 Pro28/Arg30, CDR-H2 Ser52/Gly52a/Ala52a, CDR-H4 Gln73/Phe74, were shown to occur in 24/27 (89%) HV1-69-sBnAbs that are characterized by CDR-H3-Tyr



**Figure 6. Understanding the structural role of the distinctive CDR-H2 amino acid substitutions in HV1-69-sBnAbs.** A) VDW contact analysis (black lines) shows that Ser52 of F10 and CR9114 (orange), and Ile52 of CR6261 (gray) make only intramolecular contacts; i.e., do not form contacts with their respective H5VN04s. Antibodies are shown in color; HA is in light gray. At far right, steric consequences of the germline Ile52 and the Ile52Ser substitutions are shown when the Abs are overlaid on their framework residues (RMSD ~0.5 Å). Comparing structures of the HV1-69-sBnAbs, centered on Ile52 of CR6261 (green), with F10 (yellow) and CR9114 (cyan), the Ile52Ser mutation in F10 and CR9114 enables the 2 strands to come closer together, as indicated by the yellow and cyan arrows. Distances in red indicate hypothetical steric clashes (<3 Å) that would be created if Ile52 were present in CR9114 and F10. B) Comparison between the unbound (PDB 4FQH, left) and H5VN04-bound structures (PDB 4FQI, right) of CR9114, colored according to the magnitude of structural change after superposition on the main-chain of the VH domain (from blue = 0 Å, through white = 1 Å, to red = 1.8 Å). CDRs and side-chains of the major contact residues are shown, as depicted in Figure 1A. Distances between the C<sub>α</sub> and C<sub>β</sub> atoms of Phe54 and the C<sub>α</sub> atom of CDR-H3 Tyr98 (shown as dashed lines) are indicated. Large rotations of the side chains of CDR-H3 Tyr98, CDR-H2 Phe54 and CDR-H2 Ile53 are also evident, as previously noted [7].

doi:10.1371/journal.ppat.1004103.g006

clustered at positions 97-to-99 (**Figure 2A**). Interestingly, six of the seven V-segment substitutions arise in CDR positions that are sparse in AID SHM motifs (WRCY/RGYW) (**Figure 3B–D**). Therefore, these mutations were generated by either error-prone long patch BER and MMR processes or by random SHM events. In addition, HV1-69-sBnAbs characterized by Gly52a are shown to associate with double or triple CDR-H3-Tyr and those characterized by Ala52a that are strongly associated with Arg30 and appear to represent unique molecular signatures for a subset of HV1-69-sBnAbs.

To further interrogate the distinctive HV1-69-sBnAb substitutions a semi-synthetic *IGHV1-69* library was designed aimed at bypassing SHM machinery restrictions, while maintaining low V-segment substitutions. H5/H1 panning campaigns resulted in the isolation of HV1-69-sBnAbs characterized by a low frequency of V-segment amino acid substitutions and a high frequency of CDR-H3-Tyr98/99 (**Figure S4 in Text S1**). The distinctive CDR-H2 Ser52 substitution was frequently recovered from both H5VN04 and H1CA0409 pannings (80.6% and 56.7%, respectively), and to

a lesser extent (~30%–40%) CDR-H1 Val27, CDR-H2 Met53, and Asn76 were also observed (**Figure S4 in Text S1**). The dominance of CDR-H2 Ser52 led us to investigate its functional importance. BCR and phage-Ab studies with *IGHV1-69/F10* reverted germline V-segment mutants showed the benefit of CDR-H2 Ser52 as a single mutation (**Figure 5A**) and mutagenesis studies confirm its importance in the natural HV1-69-sBnAbs (**Figure 5B–C**). Structural modeling (**Figure 6**) suggests that the CDR-H2 Ser52 and P52aG/A substitutions do not make direct contact with HA, rather they likely facilitate the ability of CDR-H2 Phe54 and CDR-H3 Tyr98 to come into close apposition and insert into adjacent pockets in the stem (**Figure 1B; Figure 6**). The findings of enriched glycine and proline residues opposed to CDR-H3-Tyr98/99, particularly for heterosubtypic Abs also adds support that CDR-H3 loop flexibility may facilitate the proper CDR-H2/H3 loop alignments for optimal binding. CR6261 lacks CDR2-H2 52/52a substitutions and uses a modified strategy which relies on distinctive substitutions occurring in the CDR-H1/H4 domain. The study of Lingwood et al [26] has shown that high H1 reactivity of CR6261 was restored only when 7 mutations, including T28P and S30R in CDR-H1 and heavily diversified CDR-H4 were back introduced into the *IGHV1-69* germline gene.

Current dogma suggests that robust B cell expansion to seasonal vaccination is often limited to B-cells that target more plentiful epitopes on the HA head domain that may outcompete for critical resources. Human immune response studies of the pdm2009 H1N1 virus have shown a relative shift in the antibody repertoire towards heterosubtypic Abs attributed to the highly divergent sequence of its head domain compared to seasonal H1N1 [11–13,34]. Isolated mAbs from these studies are characterized by high frequency of SHM suggesting that these originated from preexisting small population of heterosubtypic memory B-cells that were expanded through increased interaction with immune cells including follicular T cells (Tfh) and their soluble mediators [12,13,35–37]. In contrast to the above, our results suggest that generation of HV1-69-sBnAbs may not require prolonged recycling/reentry of B cells into germinal centers to acquire additional SHMs since both the BCR studies (**Figure 5A**) and panning studies (**Figure S4 in Text S1**) demonstrated that as few as one (**Figure 5A**) or two (**Figure S4 in Text S1**) V-segment substitutions respectively, together with properly positioned Tyr in the CDR-H3 domain can result in heterosubtypic binding and neutralization. This reasoning further suggests that HV1-69-sBnAb B-cells may be derived from a heterogeneous population of memory B cells, including B cells that exist the germinal center early without the need to extensively mutate their rearranged VH genes and IgM<sup>+</sup>CD27<sup>+</sup> marginal zone (MZ) B cells that are proposed to develop outside of the germinal center [24]. B cell helper neutrophils (N<sub>BH</sub>) have been shown to trigger SHM in MZ B cells through an extrafollicular pathway that may not require T cells [38]. In support of this proposition the Abs reported by Throsby et al [3] (**Table S1 in Text S1**) are derived from MZ B cells and HV1-69-sBnAbs CR6331/CR6342 (**Figure 3A**) contain only 5 V-segment mutations and have critical substitutions that occur in non-classical SHM positions. In addition, it appears that generation of stem targeted mAbs does not require active germinal centers as the study of Keating et al [39] showed that mice treated with rapamycin and immunized with A/HK/x31 (H3N2) were able to generate Ab response that targets the stem domain of H5VN04 HA.

While there are reports of successful vaccination approaches that lead to robust elicitation of sBnAbs in animals [40–42], it remains to be determined if these approaches will elicit a diverse pool of sBnAbs in human [43]. We propose that, it is worthwhile

to consider how to inclusively elicit HV1-69-sBnAbs in human vaccines because of their virologic potency. Why broad elicitation of naive B-cell derived HV1-69-sBnAbs was not the dominant component of the Ab response seen to pdm2009 H1N1 and how primary and recall sBnAb responses can be augmented by vaccination are questions yet to be answered. Similar to the evolving vaccination approaches that are being applied to HIV which include immunogen designs that target VRC01 precursor B-cells [44], selective clonal expansion of HV1-69-sBnAb precursor B-cells might be achieved by a stem-epitope containing immunogen [45–48] or by anti-idiotype [49] priming that results in stimulation of a larger *IGHV1-69* BCR precursor pool. In addition, the molecular signatures discovered in this study might be also useful for confirming success of such vaccination approaches in humans. For example, the Ab repertoire of individuals can be analyzed pre and post-influenza vaccination by using NGS [50]. In this regard, screening the antibodyome for the HV1-69-sBnAb molecular signatures as defined in this study would aid in quantitating their contribution to the sBnAb response. Likewise given the allelic variation of the Ig locus in the population, a complementary genomics approach should be considered to evaluate the role of *IGHV1-69* polymorphism in the sBnAb response [33,51].

## Materials and Methods

### Materials

The anti-HA antibodies F10, A66, G17, and D8 were previously described in the study of Sui et al [1]. The mAbs CR6261 and CR6331 were synthesized by Genewiz, North Brunswick, NJ. Recombinant HA of H5VN04 was produced as described [1]. A/California/04/2009 (H1CA0409) and A/Singapore/1/57 (H2SIN57) recombinant HAs were supplied by Biodefense and Emerging Infections Research Resources Repository (BEI Resources).

### Generation of antibody variants

*IGHV1-69\*01* germline V-segment was synthesized by Geneart (Regensburg, Germany). The germline variant of VH1-69/F10 was constructed by ligating the *IGHV1-69\*01* gene (NcoI 5', BssHII 3') with F10 gene segment that included the CDR-H3+light chain (BssHII 5' NotI 3') into the pET22b vector, which was digested between NcoI-NotI. The various other VH1-69/F10 variants, HV1-69-sBnAbs derivatives of F10, A66, G17, D8, CR6261, and CR6331, as well as the F10 and A66 CDR-H2 variants were constructed using QuikChange Lightning Site-Directed Mutagenesis Kit (Stratagene).

### Expression and purification of scFv

The scFv antibody sequences were cloned into the bacterial expression vector pET22b with an in-frame fusion of streptactin tag at the carboxy-terminus end. Plasmids were transformed into the expression BL21 (DE3) strain and the scFvs were produced by using the overnight express medium [52] according to the manufacturer protocol (Novagen). The scFvs were purified from clear bacterial cell lysates using the high-bind sepharose streptactin beads.

### Kinetic studies

Surface plasmon resonance (SPR) analysis was utilized for all kinetic measurements with a Biacore T100. For H5 binding kinetic studies, carboxyl terminus histidine tagged-H5 [1] was captured on a NTA-Ni<sup>+</sup> activated chip. After stabilization period of 300 sec, the scFv in question was injected by using the single

cycle kinetics function. Mobile phase contained HBS-P supplemented with 50 μM EDTA. Chip regeneration was carried out with two pulses of 0.3 M EDTA followed by injection of Nickel 50 μM of Ni<sup>+</sup> solution.

### Panning of the phage display libraries

Panning of the phage display libraries was performed by standard immunotube approach [1].

### Binding and F10-epitope mapping competition assays of phage-Abs to HA antigens as determined by MSD-based ELISA assays

The Sector Imager 2400 from Meso Scale Discovery (MSD, Rockville, MD) is utilized for interrogating the binding activities between antibody and their respective antigens based on the manufacturer's instructions. Testing of purified fd-tet derived VH1-69/F10 phage-Ab variants was carried out by spot coating 6.25 ng of purified H5VN04 HA antigens onto 384-well high-bind MSD plates followed by incubation with serially diluted phage-Ab prep in triplicates. For F10-epitope competition assay with purified anti-H5VN04/H1CA0409 phage-Abs, the phage-Abs were added in duplicates to a plate precoated with purified HA from H5VN04 or H1CA0409 or H2SIN57, and blocked with F10-scFv or an irrelevant control scFv. Phage-Ab binding was detected with Sulfo-tagged anti-M13 mAb and assayed with a MSD Sector Imager 2400.

### Phage-Ab mediated neutralization assay

Phage-Ab mediated neutralization assay with H5V04 or H1PR8 pseudotyped luciferase-reporter lentiviral particles was performed according to previous published protocol [1] using purified phage-Abs at the concentration of 1.07e13 phage particles per ml.

### B-cell activation induced by BCR cross-linking

B-cell activation induced by BCR cross-linking was performed according to the study of Hoot et al [53].

### Binding of scFvs to H5VN04 as determined by standard ELISA assay

Standard ELISA assay was used to detect binding of the scFvs to H5VN04. Briefly, 2 μg/ml of H5VN04 was coated onto 384-well plates. Upon blocking with 2%BSA, purified strep-tagged scFvs were added onto the H5-coated plates and the binding was detected with Strep-Tactin-HRP mAb conjugate (IBA, GMBH) using PolarStar at 450 nm.

### Circular dichroism measurements

Circular dichroism measurements were performed with Aviv circular dichroism spectrometer model 202 (Aviv, Lakewood, NJ). PBS was served as a blank buffer control. The experimental conditions of the CD measurements were: averaging time –5 s; wavelength steps –1 nm; range from 195 to 255 nm; and temperature 25°C during the runs.

### Structural analysis and modeling

Molecular graphics and analyses were performed using PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC, and by using the UCSF Chimera package [54] (<http://www.cgl.ucsf.edu/chimera>). Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). The in-silico mutagenesis modeling was preformed according to the study of Fahmy et al [55].

### CDR definition and antibody numbering scheme

Our approach for defining the CDR domains was based on the IMGT definitions. However, for clarity reason we have decided to use the KABAT numbering system. In addition although not formally acknowledged, we have also defined a CDR-H4 domain as detailed in supplementary **Figure S2 in Text S1**.

### Data assembly and statistical analysis

HV1-69-sBnAb sequences were obtained through the NCBI website or published patents with the exception of the Ab named Kashyap1. Kashyap et al have published 61 clonally related heterosubtypic Abs in their 2008 study [2]. From the respective patent we have chosen the sequence of the HV1-69-sBnAb designated Ab1, which is claimed to be a heterosubtypic neutralizer. The reference dataset of functional *IGHV1-69* 51p1-allele germline based Abs was constructed using the Ig Blast website ([www.ncbi.nlm.nih.gov/igblast/retrieveig.html](http://www.ncbi.nlm.nih.gov/igblast/retrieveig.html)). Default parameters were kept for the categories of length and identity, synthetic Ab sequences were excluded, and in the germline gene name category the *IGHV1-69* 51p1 allele group gene were entered as *IGHV1-69\*01, 03, 05, 06, 07, 12* and *13*.

The retrieved 7 datasets were compiled into one 51p1 allele based Ab dataset and duplicated sequences were removed. In order to obtain a dataset characterized by Abs that start with first V-segment codon of Q1 (C.A.G) and do not surpass S113 (the last amino acid of the J-segment), the dataset was mapped to the reference *IGHV1-69\*01* gene, which allowed to crop Ab sequences that start with Q1 and then the cropped dataset was mapped again against a consensus J-segment (WGQGTLVTSS) allowing the deletion of nucleotide sequences that go beyond S113 from the dataset. To facilitate the removal of clonally related Abs from the dataset, a CDR-H3 sub-alignment (C92-to-W103) was extracted and a sequence similarity matrix was organized by the name of the study. Studies found to be composed of identical CDR-H3 sequences (100% sequence identity) were taken out of the dataset. The resultant dataset was further cleaned by removal of sequence characterized by ambiguous nucleotide notations and of the studies detailed in **Table S1 in Text S1**. The entire dataset was translated, and was deleted of duplicated V-segments.

### Identification of unique amino acid substitutions in the HV1-69-sBnAb dataset

Using the UGENE software a matrix of amino acid substitutions was generated for the HV1-69-sBnAb and for the reference *IGHV1-69*-Ab datasets. A two-step method was used to identify distinctive amino acid substitutions associated with the HV1-69-sBnAb dataset. First, a Fisher's exact test was used to compare the distribution of amino acid substitutions at each position within the V-segment in the HV1-69-Abs dataset with that in the *IGHV1-69*-Ab reference dataset. Next, for germline positions where a significant statistical difference was found ( $P < 0.05$ ), another set of Fisher's exact tests were performed to compared the frequency of single amino acid substitutions. For the comparisons of individual substitution pattern at a given position, Bonferroni adjusted P-values were used to determine statistical significance in order to maintain an overall Type I error rate of 0.05 or less at each V-segment position.

### Identification of significantly enriched residues post phage-Ab selection

In order to study which of the non-germline amino acids in the semi-synthetic library were significantly enriched post H5VN04 and H1CA0409 selections (**Figure 4C and Figure S4B, E in**

**Text S1**, and which CDR-H3 amino acids were enriched in the anti-H5VN04 heterosubtypic phage-Ab pool (**Figure S4C**) a matrix of amino acid substitutions was generated as described above. The hypothesis testing procedure was also performed as described above, except that chi-squared test was used to compare the distribution of amino acid substitution at the diversified positions against their respective positions in the design scheme (**Figure S3 in Text S1**). P-values were obtained by Monte-Carlo simulations, as chi-square approximation may not be appropriate with small sample sizes.

## Supporting Information

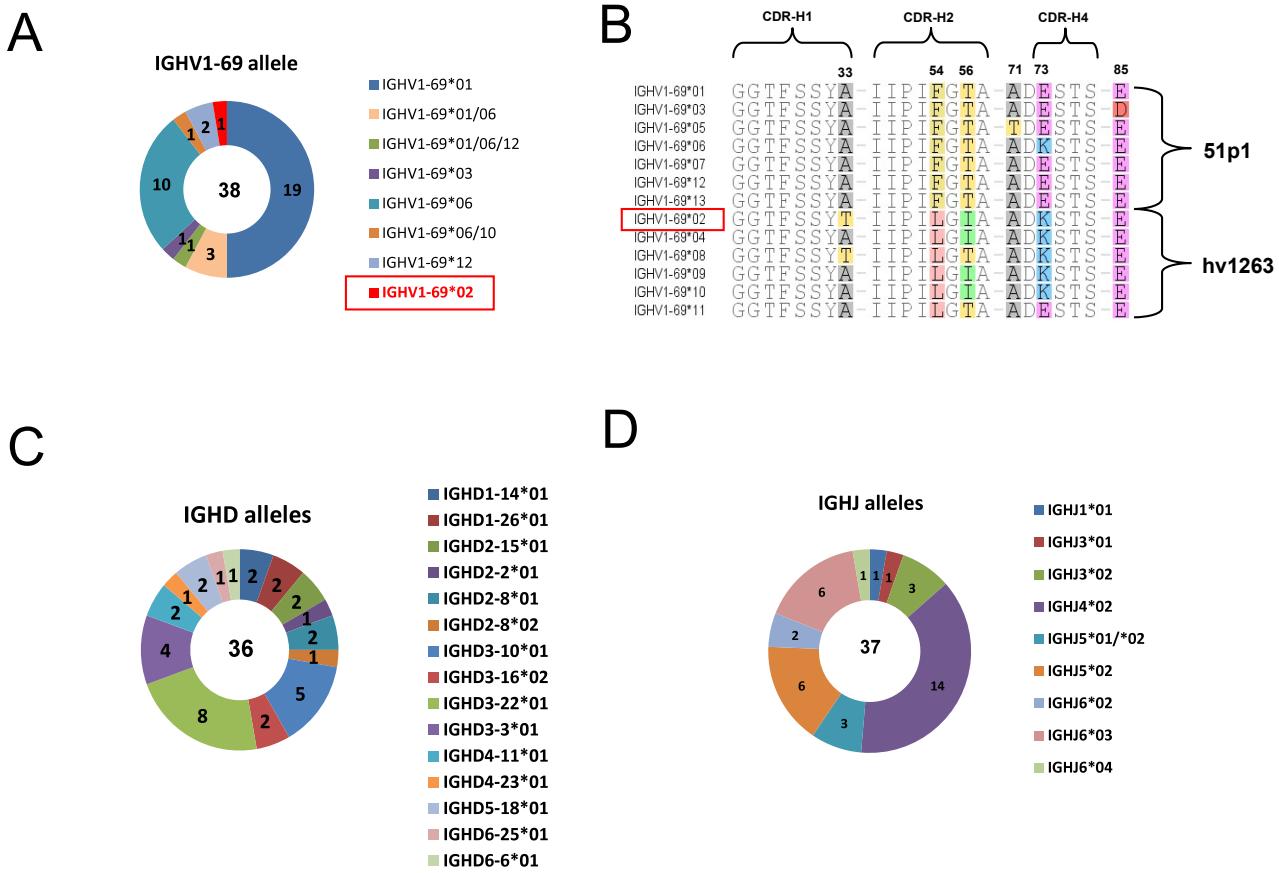
**Text S1** The Supporting Text S1 file includes: Supporting Figure S1, which details the immunogenetic analysis of HV1-69-sBnAbs; Supporting Figure S2, which details how the CDR-H4 loop was defined; Supporting Figure S3, which describes the design scheme of the semi-synthetic *IGHV1-69* Ab library; Supporting Figure S4, which details the isolated anti-H5VN04

## References

- Sui J, Hwang WC, Perez S, Wei G, Aird D, 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.
- Kashyap AK, Steel J, Oner AF, Dillon MA, Swale RE, et al. (2008) Combinatorial antibody libraries from survivors of the Turkish H5N1 avian influenza outbreak reveal virus neutralization strategies. *Proc Natl Acad Sci U S A* 105: 5986–5991.
- Throsby M, van den Brink E, Jongeneelen M, Poon LL, Alard P, et al. (2008) Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells. *PLoS One* 3: e3942.
- Ekiert DC, Bhabha G, Elsliger MA, Friesen RH, Jongeneelen M, et al. (2009) Antibody recognition of a highly conserved influenza virus epitope. *Science* 324: 246–251.
- Ekiert DC, Friesen RH, Bhabha G, Kwaks T, Jongeneelen M, et al. (2011) A highly conserved neutralizing epitope on group 2 influenza A viruses. *Science* 333: 843–850.
- Corti D, Voss J, Gamblin SJ, Codoni G, Macagno A, et al. (2011) A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* 333: 850–856.
- Dreyfus C, Laursen NS, Kwaks T, Zuidgeest D, Khayat R, et al. (2012) Highly conserved protective epitopes on influenza B viruses. *Science* 337: 1343–1348.
- Nakamura G, Chai N, Park S, Chiang N, Lin Z, et al. (2013) An in vivo human-plasmablast enrichment technique allows rapid identification of therapeutic influenza A antibodies. *Cell Host Microbe* 14: 93–103.
- Friesen RH, Lee PS, Stoop EJ, Hoffman RM, Ekiert DC, et al. (2014) A common solution to group 2 influenza virus neutralization. *Proc Natl Acad Sci U S A* 111: 445–450.
- Corti D, Suguitan AL, Jr., Pinna D, Silacci C, Fernandez-Rodriguez BM, et al. (2010) Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. *J Clin Invest* 120: 1663–1673.
- Wrammert J, Koutsanatos D, Li GM, Edupuganti S, Sui J, et al. (2011) Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. *J Exp Med* 208: 181–193.
- Thomson CA, Wang Y, Jackson LM, Olson M, Wang W, et al. (2012) Pandemic H1N1 Influenza Infection and Vaccination in Humans Induces Cross-Protective Antibodies that Target the Hemagglutinin Stem. *Front Immunol* 3: 87.
- Li GM, Chiu C, Wrammert J, McCausland M, Andrews SF, et al. (2012) Pandemic H1N1 influenza vaccine induces a recall response in humans that favors broadly cross-reactive memory B cells. *Proc Natl Acad Sci U S A* 109: 9047–9052.
- Sui J, Sheehan J, Hwang WC, Bankston LA, Burchett SK, et al. (2011) Wide prevalence of heterosubtypic broadly neutralizing human anti-influenza A antibodies. *Clin Infect Dis* 52: 1003–1009.
- Pica N, Hai R, Krammer F, Wang TT, Maamary J, et al. (2012) Hemagglutinin stalk antibodies elicited by the 2009 pandemic influenza virus as a mechanism for the extinction of seasonal H1N1 viruses. *Proc Natl Acad Sci U S A* 109: 2573–2578.
- Miller MS, Tsibane T, Krammer F, Hai R, Rahmat S, et al. (2013) 1976 and 2009 H1N1 influenza virus vaccines boost anti-hemagglutinin stalk antibodies in humans. *J Infect Dis* 207: 98–105.
- Boyd SD, Gaeta BA, Jackson KJ, Fire AZ, Marshall EL, et al. (2010) Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J Immunol* 184: 6986–6992.
- Glanville J, Kuo TC, von Budingen HC, Guey L, Berka J, et al. (2011) Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. *Proc Natl Acad Sci U S A* 108: 20066–20071.
- Schroeder HW, Jr., Hillson JL, Perlmann RM (1987) Early restriction of the human antibody repertoire. *Science* 238: 791–793.
- Sasso EH, Willems van Dijk K, Bull AP, Milner EC (1993) A fatally expressed immunoglobulin VH1 gene belongs to a complex set of alleles. *J Clin Invest* 91: 2358–2367.
- Milner EC, Hufnagle WO, Glas AM, Suzuki I, Alexander C (1995) Polymorphism and utilization of human VH Genes. *Ann N Y Acad Sci* 764: 50–61.
- Harris AK, Meyerson JR, Matsuo Y, Kuybeda O, Moran A, et al. (2013) Structure and accessibility of HA trimers on intact 2009 H1N1 pandemic influenza virus to stem region-specific neutralizing antibodies. *Proc Natl Acad Sci U S A* 110: 4592–4597.
- Victora GD, Nussenzweig MC (2012) Germinal centers. *Annu Rev Immunol* 30: 429–457.
- Cerutti A, Cols M, Puga I (2013) Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes. *Nat Rev Immunol* 13: 118–132.
- Meireles LM, Domling AS, Camacho CJ (2010) ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery. *Nucleic Acids Res* 38: W407–411.
- Lingwood D, McTamney PM, Yassine HM, Whittle JR, Guo X, et al. (2012) Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature* 489: 566–570.
- Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, et al. (2011) Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* 333: 1593–1602.
- Peled JU, Kuang FL, Iglesias-Ussel MD, Roa S, Kalis SL, et al. (2008) The biochemistry of somatic hypermutation. *Annu Rev Immunol* 26: 481–511.
- Rogozin IB, Pavlov YI, Bebenek K, Matsuda T, Kunkel TA (2001) Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. *Nat Immunol* 2: 530–536.
- Clark LA, Ganesan S, Papp S, van Vlijmen HW (2006) Trends in antibody sequence changes during the somatic hypermutation process. *J Immunol* 177: 333–340.
- Almagro JC, Beavers MP, Hernandez-Guzman F, Maier J, Shaulsky J, et al. (2011) Antibody modeling assessment. *Proteins* 79: 3050–3066.
- Corti D, Lanzavecchia A (2013) Broadly neutralizing antiviral antibodies. *Annu Rev Immunol* 31: 705–742.
- Sasso EH, Johnson T, Kipps TJ (1996) Expression of the immunoglobulin VH gene 51p1 is proportional to its germline gene copy number. *J Clin Invest* 97: 2074–2080.
- Han T, Marasco WA (2011) Structural basis of influenza virus neutralization. *Ann N Y Acad Sci* 1217: 178–190.
- Kuchen S, Robbins R, Sims GP, Sheng C, Phillips TM, et al. (2007) Essential role of IL-21 in B cell activation, expansion, and plasma cell generation during CD4+ T cell-B cell collaboration. *J Immunol* 179: 5886–5896.
- Vinuesa CG, Linterman MA, Goodnow CC, Randall KL (2010) T cells and follicular dendritic cells in germinal center B-cell formation and selection. *Immunol Rev* 237: 72–89.
- Schwickert TA, Victora GD, Fooksman DR, Kamphorst AO, Mugnier MR, et al. (2011) A dynamic T cell-limited checkpoint regulates affinity-dependent B cell entry into the germinal center. *J Exp Med* 208: 1243–1252.

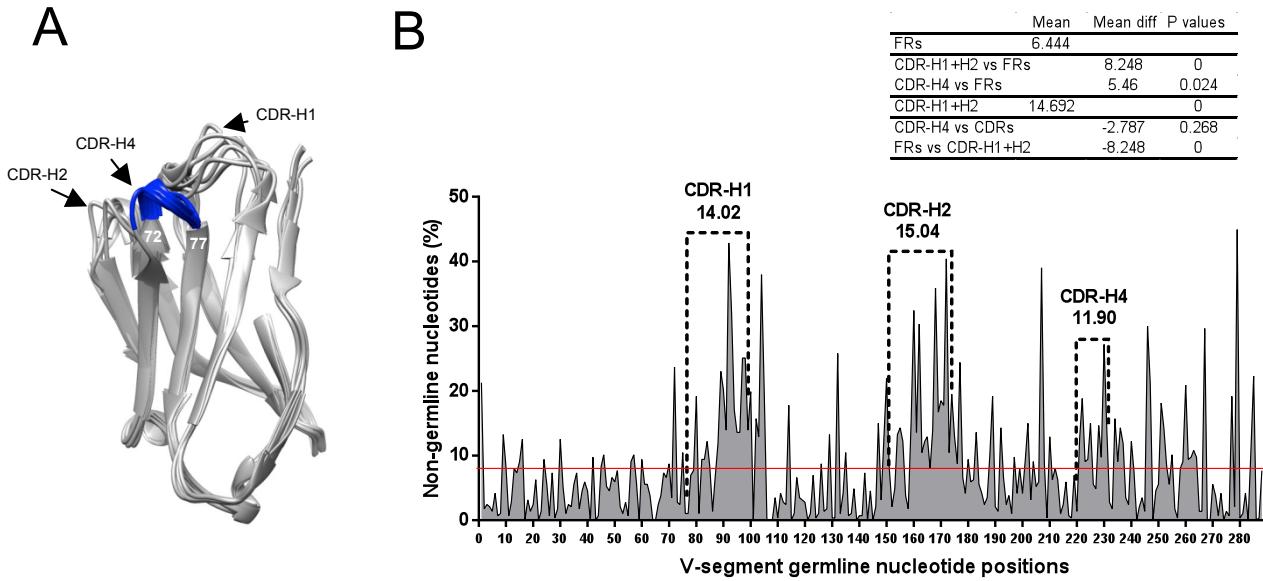
38. Puga I, Cols M, Barra CM, He B, Cassis L, et al. (2012) B cell-helper neutrophils stimulate the diversification and production of immunoglobulin in the marginal zone of the spleen. *Nat Immunol* 13: 170–180.
39. Keating R, Hertz T, Wehenkel M, Harris TL, Edwards BA, et al. (2013) The kinase mTOR modulates the antibody response to provide cross-protective immunity to lethal infection with influenza virus. *Nat Immunol* 14: 1266–1276.
40. Wei CJ, Yassine HM, McTamney PM, Gall JG, Whittle JR, et al. (2012) Elicitation of broadly neutralizing influenza antibodies in animals with previous influenza exposure. *Sci Transl Med* 4: 147ra114.
41. Goff PH, Eggink D, Seibert CW, Hai R, Martinez-Gil L, et al. (2013) Adjuvants and immunization strategies to induce influenza virus hemagglutinin stalk antibodies. *PLoS One* 8: e79194.
42. Eggink D, Goff PH, Palese P (2014) Guiding the Immune Response against Influenza Virus Hemagglutinin toward the Conserved Stalk Domain by Hyperglycosylation of the Globular Head Domain. *J Virol* 88: 699–704.
43. Krammer F, Palese P (2013) Universal influenza virus vaccines: need for clinical trials. *Nat Immunol* 15: 3–5.
44. Jardine J, Julien JP, Menis S, Ota T, Kalyuzhnii O, et al. (2013) Rational HIV immunogen design to target specific germline B cell receptors. *Science* 340: 711–716.
45. Krammer F, Pica N, Hai R, Margine I, Palese P (2013) Chimeric hemagglutinin influenza virus vaccine constructs elicit broadly protective stalk-specific antibodies. *J Virol* 87: 6542–6550.
46. Steel J, Lowen AC, Wang TT, Yondola M, Gao Q, et al. (2010) Influenza virus vaccine based on the conserved hemagglutinin stalk domain. *MBio* 1: e00018–10.
47. Bommakanti G, Lu X, Citron MP, Najar TA, Heidecker GJ, et al. (2012) Design of Escherichia coli-expressed stalk domain immunogens of H1N1 hemagglutinin that protect mice from lethal challenge. *J Virol* 86: 13434–13444.
48. Lu Y, Welsh JP, Swartz JR (2014) Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines. *Proc Natl Acad Sci U S A* 111: 125–130.
49. Potter KN, Li Y, Mageed RA, Jefferis R, Capra JD (1999) Molecular characterization of the VH1-specific variable region determinants recognized by anti-idiotypic monoclonal antibodies G6 and G8. *Scand J Immunol* 50: 14–20.
50. Finn JA, Crowe JE, Jr. (2013) Impact of new sequencing technologies on studies of the human B cell repertoire. *Curr Opin Immunol* 25: 613–618.
51. Watson CT, Steinberg KM, Huddleston J, Warren RL, Malig M, et al. (2013) Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. *Am J Hum Genet* 92: 530–546.
52. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41: 207–234.
53. Hoot S, McGuire AT, Cohen KW, Strong RK, Hangartner L, et al. (2013) Recombinant HIV Envelope Proteins Fail to Engage Germline Versions of Anti-CD4bs bNAbs. *PLoS Pathog* 9: e1003106.
54. Petersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612.
55. Fahmy A, Wagner G (2011) Optimization of van der Waals energy for protein side-chain placement and design. *Biophys J* 101: 1690–1698.

## Supporting Text 1



**Figure S1. Immunogenetic analysis of HV1-69-sBnAbs.**

HV1-69-sBnAbs were analyzed for **A)** V-segment allele usage ( $n = 38$ ), with panel **B)** showing the 13 known *IGHV1-69* alleles and their classification into the 51p1 and hv1263 allele groups. Panel **C)** shows D-segment usage ( $n = 36$ ), and panel **D)** shows J-segment usage ( $n = 37$ ).



**Figure S2. Defining the CDR-H4 loop in *IGHVI-69* 51p1 allele based Abs.**

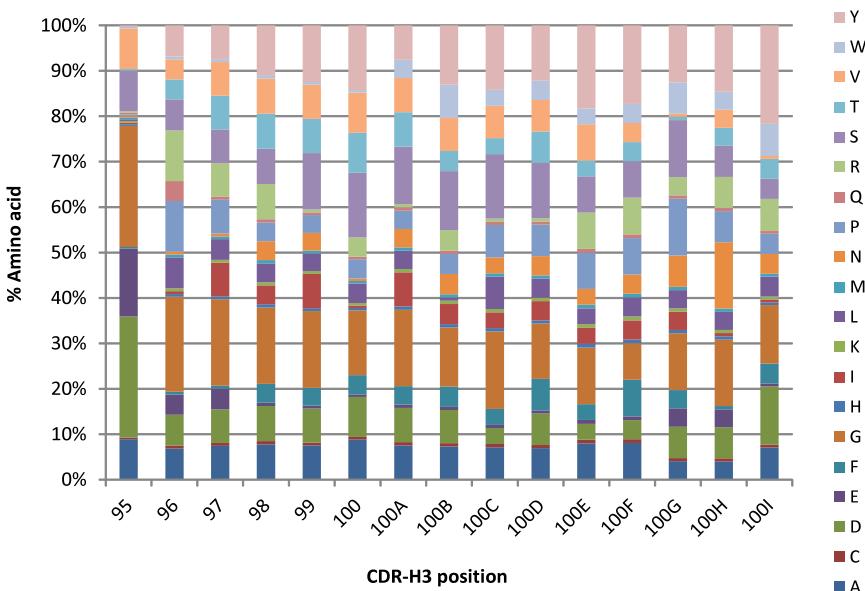
The existence of a CDR-H4 loop, or hypervariable loop-4 (HV4), has been suggested by several studies [1,2]. However, no formal definition has been given to this loop. In this study the approach for defining the CDR-H4 loop was based on structural alignment, and by studying the overall nucleotide substitution frequency of the structurally defined CDR-H4 region as it appears in the reference *IGHVI-69*-Ab 51p1 allele based dataset (sees Methods for more details). **(A)** Structural alignment was performed for 7 non-antigen complexed *IGHVI-69* 51p1 allele related Abs: E51, 47e, 412d, CR9114, 1-69/b3, CR6261 and N12-i2. With the exception of N12-i2 all other six Abs were characterized by a loop that starts with position 73 and ends with position 76. Accordingly, the germline CDR-H4 loop sequence is defined as E.S.T.S. **(B)** The *IGHVI-69*-Ab 51p1 allele based reference dataset was analyzed for non-germline nucleotide substitution frequencies. The red line shows the mean of non-germline nucleotide substitution frequency observed for FR regions and dashed lines point to the CDR areas. The analysis of variance (ANOVA) shows **(B — inset)** that the mean of non-germline nucleotide substitution frequency of the CDR-H4 area is significantly higher than that of the FRs ( $P = 0.02$ ), but is not significantly different than that of the CDR-H1+H2 ( $P = 0.27$ ).

A

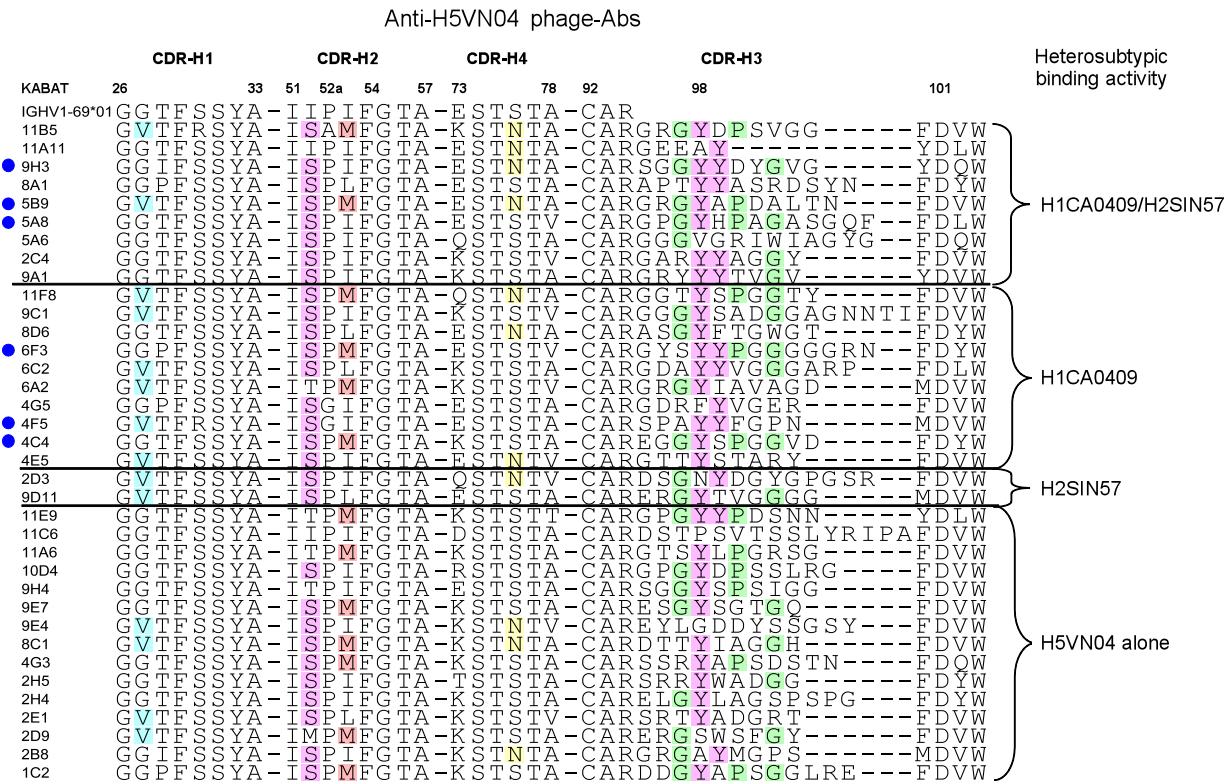
CDR-H1			CDR-H2				CDR-H4				J		
	Gly27	Thr28	Ser30	Ile52	Pro52a	Ile53	Phe54	Glu73	Ser74	Ser76	Ala78	PheF100	Val102
A				1.4%	10.0%				2.3%		81.68%		
C													
D								2.3%					
E							60.5%						
F					4.6%	95.0%			5.0%			70.0%	
G	90.0%				10.0%					2.5%			
H													
I		10.0%		71.1%		53.3%							
K								27.8%					
L				1.4%		8.15%	5.0%					13.3%	
M				1.4%		10.6%						15.0%	
N				1.4%		8.2%				11.0%			
P		10.0%			71.6%			5.0%				13.3%	
Q								2.3%		5.4%			
R			10.0%					92.7%	75.7%	1.8%			
S			90.0%	10.0%	1.1%	2.3%							
T			80.0%	8.4%	7.2%	2.3%		2.3%		5.4%	1.8%		
V		10.0%			4.8%		10.6%				14.7%		60.0%
W													
Y												15.0%	13.3%

Values are rounded to the nearest tenth.

B

**Figure S3. The design principles of the *IGHV1-69* synthetic library.**

**A)** Diversification plan of the V and J segments. In red are amino acids that were elevated beyond their natural observed frequency. **B)** Diversification scheme for the CDR-H3 domain, which was based on the natural frequency and diversity observed in a reference CDR-H3 alignment made from *IGHV1-69* 51p1allele based Abs (n=1217).

**A****B**

Library design scheme	Anti-H5VN04 phage-Abs	Ratio	P <sup>a</sup>
HV1-69-sBnAb distinctive amino acid substitutions*			
<b>CDR-H1</b>			
G27V	10%	39%	3.9
<b>CDR-H2</b>			
Ser52	10%	81%	8.1
Non-HV1-69-sBnAb distinctive amino acid substitutions			
<b>CDR-H2</b>			
I53M	11%	36%	3.4
<b>CDR-H4</b>			
S76N	11%	31%	2.8
<b>CDR-H3</b>			
G97	19%	53%	2.8
Y98	11%	78%	7.1
Y99	12%	33%	2.7
P100	4%	34%	8.1
G100B	13%	43%	3.3

a - P-values were obtained by Monte-Carlo simulations.

\*- See Figure 2B.

# C

Library design scheme	Heterosubtypic subset CDR-H3 (n = 16)	Non-heterosubtypic subset CDR-H3 (n = 8)
Y98	11%	100% (16) *
Y99	12%	44%*(7)*
P100	4%	38%* (6)*
G100B	13%	63%*(10)*

\* - Significant change from the library design scheme. P < 0.05 as determined by Monte-Carlo simulations.

† Non-significant change from the library design scheme.

D

## Anti-H1CA0409 phage-Abs

	CDR-H1	CDR-H2	CDR-H4	CDR-H3	Heterosubtypic binding activity
KABAT	26	33 51 52a 54	57 73	78 92 98	101
IGHV1-69*01	GGTFSSSYA	-ITPIFGTA	-ESTSTA	-CAR	
● 9H3	GGI <sup>T</sup> FSSSYA	-IS <sup>P</sup> I <sup>T</sup> FGTA	-ESTNTA	-CARSGS <sup>G</sup> Y <sup>Y</sup> DYGVG	-YDQW
● 5B9	GGV <sup>T</sup> TFSSSYA	-IS <sup>P</sup> M <sup>T</sup> FGTA	-ESTNTA	-CARGRG <sup>Y</sup> APDALTN	-FDVW
● 5A8	GGT <sup>T</sup> FSSSYA	-IS <sup>P</sup> I <sup>T</sup> FGTA	-ESTSTV	-CARGPGYHPAGASQOF	-FDLW
2E12	GV <sup>T</sup> TFSSSYA	-IS <sup>P</sup> M <sup>T</sup> FGTA	-ESTNTA	-CARGGG <sup>G</sup> Y <sup>Y</sup> PAGVGR	-YDVW
1H2	GGP <sup>T</sup> FRSSSYA	-IS <sup>P</sup> I <sup>T</sup> FGTA	-ESTSTA	-CARGG <sup>G</sup> Y <sup>S</sup>	-FDVW
1F5	GGT <sup>T</sup> FSSSYA	-IS <sup>P</sup> I <sup>T</sup> FGTA	-ESTSTA	-CARDQ <sup>G</sup> GT <sup>R</sup> GNY	-FDVW
1F3	GV <sup>T</sup> IFSSSYA	-IS <sup>P</sup> M <sup>T</sup> FGTA	-ESTSTA	-CARSRG <sup>Y</sup> Y <sup>A</sup> PTSFH	-YDVW
1F1	GGT <sup>T</sup> FSSSYA	-IS <sup>P</sup> I <sup>T</sup> FGTA	-ESTRTA	-CARGSG <sup>G</sup> Y <sup>V</sup> AA <sup>S</sup> GA	-FDVW
1E6	GV <sup>T</sup> TFSSSYA	-IS <sup>P</sup> I <sup>T</sup> FGTA	-ESTSTA	-CARGREYYASNGDS	-FDVW
● 6F3	GGP <sup>T</sup> FRSSSYA	-IS <sup>P</sup> M <sup>T</sup> FGTA	-ESTSTV	-CARGYSY <sup>Y</sup> YPGGGGGRN	-FDYW
● 4F5	GV <sup>T</sup> TFRS <sup>S</sup> YA	-IS <sup>P</sup> I <sup>T</sup> FGTA	-ESTSTA	-CARSPAY <sup>Y</sup> FGPN	-MDVW
● 4C4	GGT <sup>T</sup> FSSSYA	-IS <sup>P</sup> M <sup>T</sup> FGTA	-KSTSTA	-CAREG <sup>G</sup> YSPGGVD	-FDYW
2G4	GGT <sup>T</sup> FSSSYA	-IS <sup>P</sup> M <sup>T</sup> FGTA	-ESTSTV	-CARSRG <sup>Y</sup> YNVAASFG	-FDVW
2E1	GGI <sup>T</sup> FRSSSYA	-IS <sup>P</sup> I <sup>T</sup> FGTA	-ESTTTV	-CARGAG <sup>S</sup> T	-FDVW
2B6	GGP <sup>T</sup> FRSSSYA	-IS <sup>P</sup> I <sup>T</sup> FGTA	-KSTNTA	-CARGRSY <sup>I</sup> IVSVSPG	-FDVW
● 2A12	GGT <sup>T</sup> FSSSYA	-IPL <sup>T</sup> FGTA	-ESTSTA	-CARGLG <sup>L</sup> Y	-FDVW
2A11	GV <sup>T</sup> IFSSSYA	-IS <sup>P</sup> V <sup>T</sup> FGTA	-DSTNTA	-CARSRG <sup>Y</sup> TVSSLAGRY	-FDQW
2A1	GV <sup>T</sup> IFSSSYA	-IPI <sup>T</sup> FGTA	-KSTSTA	-CARGGS <sup>G</sup> S	-FDVW
1H8	GV <sup>T</sup> TFSSSYA	-IS <sup>P</sup> M <sup>T</sup> FGTA	-ESTNTV	-CARGRAY <sup>L</sup> SVRG <sup>S</sup>	-FDVW
1H4	GV <sup>T</sup> TFSSSYA	-IS <sup>P</sup> I <sup>T</sup> FGTA	-QSTSTA	-CARGLG <sup>T</sup> YSPSLYPRG	-MDVW
2F1	GGTFSSSYA	-ITPI <sup>T</sup> FGTA	-DSTSTA	-CARGPTLY <sup>S</sup> PPV	-FDVW
2H4	GV <sup>T</sup> TFSSSYA	-ITPI <sup>T</sup> FGTA	-QSTSTA	-CARGG <sup>G</sup> V	-FDVW
2H3	GGP <sup>T</sup> FRSSSYA	-IPI <sup>T</sup> FGTA	-ESTSTA	-CARGTD <sup>D</sup> Y <sup>S</sup> GYRG	-FDVW
2G3	GGT <sup>T</sup> FRSSSYA	-IMPI <sup>T</sup> FGTA	-ESTTTV	-CARGAG <sup>G</sup> V <sup>S</sup> AGPSWP	-FDVW
2E11	GGT <sup>T</sup> FSSSYA	-INPI <sup>T</sup> FGTA	-ESTNTA	-CARGESAY <sup>S</sup> YSRN	-YDVW
2C1	GGI <sup>T</sup> FRSSSYA	-IS <sup>P</sup> I <sup>T</sup> FGTA	-ESTSTA	-CARDSGIA <sup>S</sup> GYTAY	-MDYW
1F12	GGI <sup>T</sup> FRSSSYA	-ITPI <sup>T</sup> FGTA	-ESTRTA	-CARDLSRD <sup>S</sup> LNLPG <sup>S</sup> SPG <sup>Y</sup> YDLW	
1F6	GGT <sup>T</sup> FRSSSYA	-ITTI <sup>T</sup> FGTA	-ESTNTA	-CARGGG <sup>G</sup> GR	-FDVW
1D9	GGT <sup>T</sup> FRSSSYA	-ITTI <sup>T</sup> FGTA	-KSTSTA	-CARGATATGF	-YDVW
● 2B11	GGT <sup>T</sup> FRSSSYA	-IPI <sup>T</sup> FGTA	-ESTNTA	-CARVR <sup>G</sup> GY <sup>P</sup> YG <sup>D</sup>	-FDVW

E

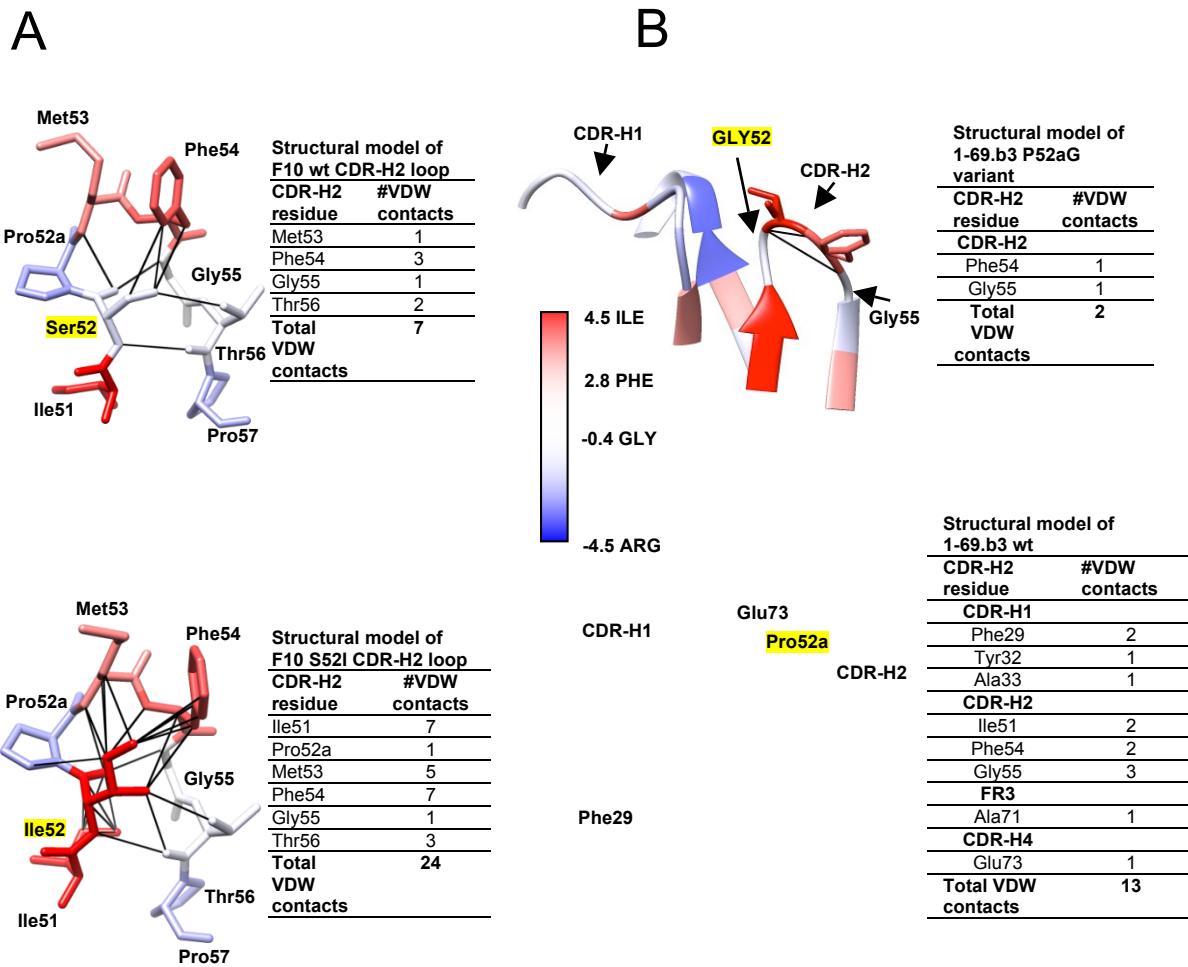
Library design scheme	Anti-H1CA0409 phage-Abs	Ratio	P <sup>a</sup>	
HV1-69-sBnAb distinctive amino acid substitutions*				
<b>CDR-H1</b>				
G27V	10%	33%	3.3	0.002
<b>CDR-H2</b>				
Ser52	10%	57%	5.7	0.0035
Non-HV1-69-sBnAb distinctive amino acid substitutions				
<b>CDR-H4</b>				
S76N	11%	30%	2.7	0.014
<b>CDR-H3</b>				
G97	19%	60%	3.1	0.01
Y98	11%	53%	4.9	0.01

a - P-values were obtained by Monte-Carlo simulations.

\*- See Figure 2B.

**Figure S4. CDR sequences and binding characteristics of the anti-H5VN04 and anti-H1CA0409 phage-Ab pools**

**A)** CDR sequence alignment of the anti-H5VN04 phage-Ab pool was ordered based on heterosubtypic binding activity as tested against H1CA0409 and H2SIN57. Highlights point to the statistically significant enriched residues post selection as discovered by Chi square test in **B**). **C)** Heterosubtypic (n=16) and non-heterosubtypic (n=8) anti-H5VN04 phage-Abs that are characterized by Ser52/Tyr98 were analyzed separately by Chi square test for the frequency of enriched amino acids in CDR-H3. **D)** CDR sequence alignment of the anti-H1CA0409 phage-Ab pool was ordered based on heterosubtypic binding activity as tested against H5VN04 and H2SIN57. Highlights point to the statistically significant enriched residues post selection as discovered by Chi square test in **E**). Tyr99 was also highlighted. **Red** circles point to the two phage-Abs which did not bind to the stem domain of H1CA0409. **Blue** circles point to common phage-Ab clones discovered independently from both panning campaigns.



**Figure S5. Understanding the structural role of the HV1-69-sBnAbs distinctive AA substitutions in positions 52 and 52a.**

**A)** In the model of non-HA complexed F10, VDW contacts were analyzed for Ser52 against other CDR-H2 loop residues (*upper panel*). Ser52 was in-silico mutagenized [3] to Ile52 to show the occurrence of a much higher number of VDW contacts (*lower panel*). **B)** *Upper panel* – Pro52a in 1-69/B3 (mAb characterized by a non-mutated *IGHVI-69\*01* V-segment [4]) was in-silico mutagenized to Gly52A to show the occurrence of minimal number of VDW contacts as compared to the germline Pro52a as shown in the *lower panel*. Coloring scheme is based on the Kyte-Doolittle hydrophobic scale.

## Supplemental Table

Study†	Panning method	Panning target	# of Positive clones <sup>a</sup>	Frequency of <i>IGHV1-69</i> Abs	# of anti-stem binding clones	Frequency of <i>IGHV1-69</i> based anti-stem Abs
Kashyap et al. (2008)	Phage display library generated from bone marrow B cells of H5N1 Turkish survivors.	H5N1		The study reports 61 clonally related <i>IGHV1-69</i> based sequences. However, the full binding properties of isolated Abs was not reported.		
Throsby et al. (2008)	Phage display library generated from IgM+ memory B cells of H5N1 vaccinated donors.	H5N1	43	15/43	13	12/13 (92%)
Sui et al. (2009)	Naïve human antibody phage display library.	H5N1	6	5/6	6	5/6 (83%)
Corti et al. (2010)+	Panning B-cells from donors vaccinated with the seasonal influenza vaccine.	H5N1/H1N1	20	14/20	19	14/19 (74%)
Wrammert et al. (2011)+	Panning plasmablasts from 2009 pandemic H1N1 infected donors.	H1N1	15	4/15	5	4/5 (80%)
Thomson et al. (2012)+	Panning B-cells and plasmablasts from 2009 pandemic H1N1 infected and vaccinated donors.	H1N1	48	25/48	7*	7/7 (100%)
Li et al. (2012) +	Panning plasmablasts from 2009 pandemic H1N1 vaccinated donors.	H1N1	28	1	3	1/3 (33%)

a — number represents distinctive VH sequence.

\*Not all positive clones were epitope mapped.

† Dreyfus et al. (2012) has reported the isolation of one *IGHV1-69* germline based anti-stem Ab.

+ Naturally paired VH and VL chains show biased *IGHV1-69* usage and promiscuous use of VL, similarly to sBnAbs isolated from Ab-phage display libraries.

**Table S1. Summary of various studies that have analyzed antibody responses towards hemagglutinin of group 1 influenza A subtypes.**

## **Supporting text:**

### **The design principles of the semi-synthetic *IGHVI-69* Ab library.**

The main goal of the *semi-synthetic IGHVI-69 Ab library* diversification scheme was to obtain an Ab library characterized by the V-segment molecular determinants associated with HV1-69-sBnAbs, while maintaining overall low V-segment amino acid substitution frequency. At the time the library was designed our knowledge of V-segment molecular determinants of *IGHVI-69*-sBnAbs was based on the structural analysis of F10 and CR6261 solved structures, and from an alignment made of HV1-69-sBnAbs reported by Thorsby et al., (2008) [5] Sui et al.,[6] (2009) and Corti et al (2010)[7]. This analysis has suggested that the amino acid substitutions of CDR-H1: Val27, Ile28, Pro29, Arg30; CDR-H2: Ser52, Gly52a, Ala52a; and CDR-H4 Glu73 and Phe74 were distinctly associated with HV1-69-sBnAbs. It was therefore decided that these amino acid substitutions should be included in the library diversification scheme at the relatively elevated frequencies of 5-to-10% (**Figure S3A**).

In order to explore if generation of HV1-69-sBnAbs based on amino acid substitutions occurring in positions CDR-H2 52 and 52a is restricted to SER52, GLY52a and ALA52a, these positions were also diversified with naturally occurring amino acid substitutions using frequencies that were acquired from a CDR-H2 alignment composed of *IGHVI-69* 51p1 allele based Abs that were devoid of germline CDR-H2 sequences (n = 800). In addition, CDR-H2 position 53 was also diversified according to this alignment for the reason that HV1-69-sBnAbs characterized by I52aG were always found to be accompanied by mutated positions 53 (either I53V or I53M) suggesting the existence of structural dependency. Likewise, since it was noticed that several HV1-69-sBnAbs are highly diversified in the surrounding CDR-H4 domain area (**Figure S2**), we decided that positions of 73, 74, 76 and 78 will be diversified with naturally occurring amino acid substitutions at a frequencies that were observed in a 73-to-78 alignment made from *IGHVI-69* 51p1 allele based Abs (n = 1477). To insure that the library would maintain “germline”

characteristics, the frequency of the respective germline amino acids in each of the diversified positions was maintained at high frequencies of 72%-to-95%.

For the CDR-H3 domain, the strategy was to have an equal presentation of CDR-H3 lengths of 5, 7, 9, 10, 11, 12, 13, and 15, whereby diversification scheme (**Figure S3B**) was based on the natural diversity observed in alignment composed of 1217 non-duplicated *IGHV1-69* 51p1 allele based Abs. In order to avoid skewing of the observed CDR-H3 frequencies due to residues associated with J-segments, the CDR-H3 sequences were aligned and deleted of amino acids post position 100 (which is Phe in IGHJ1-5). Additionally, the CDR-H3 sequences were also omitted of IGHJ6's quintuple tyrosine residues (YYYYYGYMDVWGQGTTVTVSS) as these were found to distort the amino acid frequency and diversity in the alignment. The design of the J-segment was based on a consensus sequence of: FDVWGQGTLVTVSS, in which F and V were diversified according to the frequency scheme presented in **Figure S3A**.

The library light chains were obtained from the previously constructed Mehta I/II naïve human light chain Ab libraries and were linked to the VH synthetic library via a (GLY<sub>3</sub>SER)<sub>4</sub> linker. Synthesis of the *IGHV1-69* library was performed by Morphosys GmbH based on the technology developed by Sloning GmbH. Generation of fd-tet scFv phage display library was performed as described previously [8]. Phage library size consisted of 7.7X10<sup>8</sup> members. Randomly sequencing of 164 Ab library members validated a similar diversification scheme (data not shown) and confirmed low V-segment amino acid diversity with a mean of 1.9±1.1 amino acid substitutions per V-segment.

## References:

1. Fanning SW, Horn JR (2011) An anti-hapten camelid antibody reveals a cryptic binding site with significant energetic contributions from a nonhypervariable loop. *Protein Sci* 20: 1196-1207.
2. Briney BS, Willis JR, Crowe JE, Jr. (2012) Location and length distribution of somatic hypermutation-associated DNA insertions and deletions reveals regions of antibody structural plasticity. *Genes Immun* 13: 523-529.
3. Fahmy A, Wagner G (2011) Optimization of van der Waals energy for protein side-chain placement and design. *Biophys J* 101: 1690-1698.
4. Almagro JC, Beavers MP, Hernandez-Guzman F, Maier J, Shaulsky J, et al. (2011) Antibody modeling assessment. *Proteins* 79: 3050-3066.
5. Throsby M, van den Brink E, Jongeneelen M, Poon LL, Alard P, et al. (2008) Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM<sup>+</sup> memory B cells. *PLoS One* 3: e3942.
6. Sui J, Hwang WC, Perez S, Wei G, Aird D, 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.
7. Corti D, Sugitan AL, Jr., Pinna D, Silacci C, Fernandez-Rodriguez BM, et al. (2010) Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. *J Clin Invest* 120: 1663-1673.
8. O'Connell D, Becerril B, Roy-Burman A, Daws M, Marks JD (2002) Phage versus phagemid libraries for generation of human monoclonal antibodies. *J Mol Biol* 321: 49-56.