Supporting Information

Structure-guided molecular grafting of a complex broadly neutralizing viral epitope

Goran Bajic[1,2], Max J. Maron[3], Timothy M. Caradonna[1,3], Ming Tian[4], Adam Mermelstein[5], Daniela Fera[5], Garnett Kelsoe[6,7], Masayuki Kuraoka[7] and Aaron G. Schmidt[3,8*]

[1]Laboratory of Molecular Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

[2]Department of Pediatrics, Harvard Medical School, Boston, MA, 02115, USA

[3]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, 02139, USA

[4]Program in Cellular and Molecular Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA.

[5]Department of Chemistry and Biochemistry, Swarthmore College, Swarthmore, PA 19081, USA

[6]Duke Human Vaccine Institute, Duke University, Durham, NC 27710, USA

[7]Department of Immunology, Duke University, Durham, NC 27710, USA

[8]Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA

***Correspondence:**
Aaron G. Schmidt
Tel: 857-268-7118; E-mail: aschmidt@crystal.harvard.edu

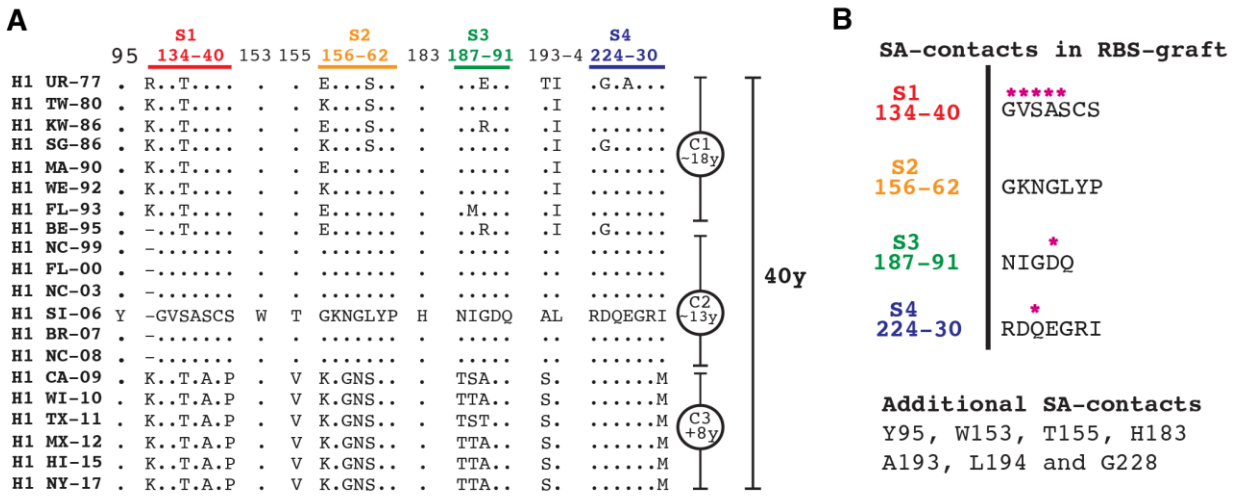


Figure S1: Conservation of the H1 HA RBS and critical SA contacts. A, Sequence alignment of historical H1 RBS and critical residues comprising sialic acid (SA) contacts. The segments in the RBS graft are colored. Conserved residues (.) are in reference to H1 SI-06. The representative antigenic clusters (C1, C2, C3) of H1 isolates are listed with the numbering of circulating corresponding years (y). In order, the representative H1 isolates are: USSR-1977, Taiwan-1980, Singapore-1986, Massachusetts-1990, Wellington-1990, Florida-1993, Bejing-1999, Florida-1990, North Carolina-2003, Solomon Islands-2006, Brisbane-2007, North Carolina-2008, California-2009, Wisconsin-2010, Texas-2010, Mexico-2012, Hawai'i-2015, New York-2017. **B,** SA-contact residues in the RBS-graft segments are noted as asterisks (magenta) with the additional, conserved SA-contacts not present in the graft are listed.

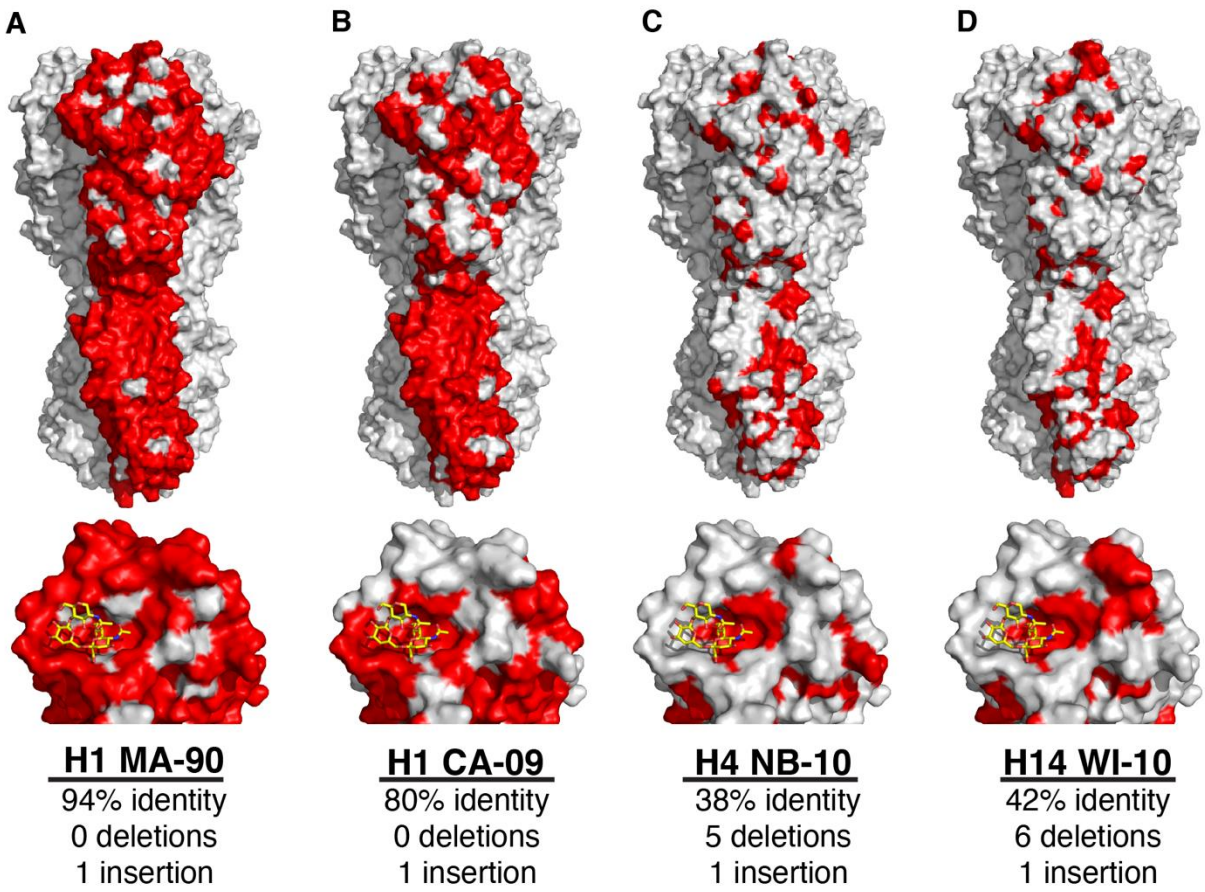


Figure S2: Conservation of HAs. Using H1 SI-06 as reference, residue conservation (sequence identity) is shown in red for historical H1 MA-90 and the new pandemic, H1 California/04/2009 (H1 CA-09) as well as the two acceptor scaffolds H4 NB-10 and H14 WI-10. Two views are shown: top is the HA trimer in space filling model with only one monomer colored red at points of conservation; the bottom is a close-up the RBS with a LStc molecule (stick-representation) docked for point of reference. The percent identity is in reference to H1-SI-06 with the total number of insertions and deletions listed. For both H1 MA-90 and H1 CA-09, the insertion is residue K133a.

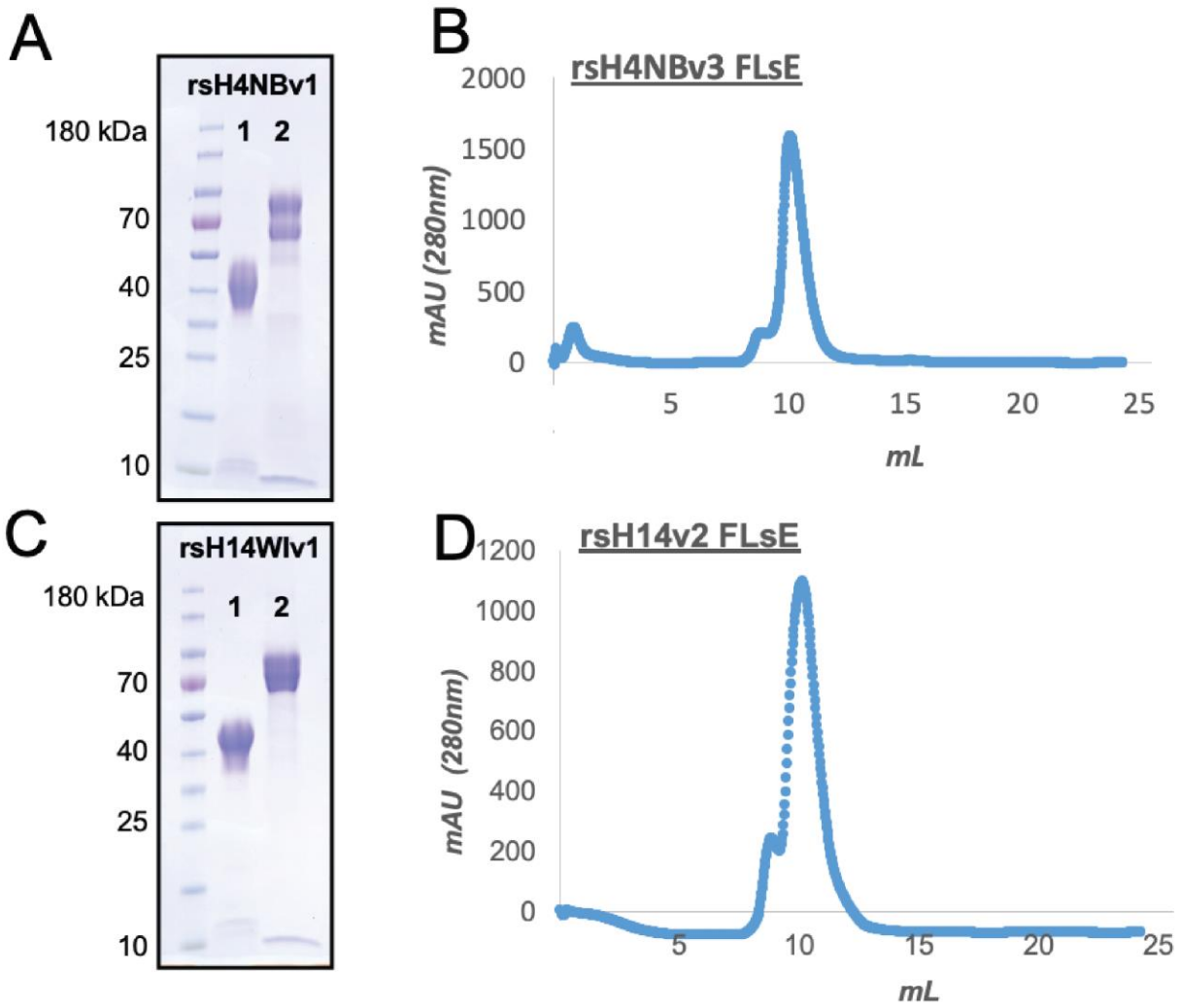


Figure S3: Biochemical characterization of the optimized rsHAs. A, Coomassie-stained SDS-PAGE gel of the rsH4NBv1 head and FLsE (full-length soluble ectodomain) constructs (marked "1" and "2", respectively). A prestained protein ladder is in the first and the corresponding molecular weights (in kilodaltons) are marked. **B,** Representative FPLC trace using a Superdex 10/300 column of the FLsE construct. The trace monitors the absorbance (in mAU) at 280nm as a function of elution volume (mL). **C,** Coomassie-stained SDS-PAGE of the rsH14WIv1 constructs (labeled as in **A)**) and **D,** representative FPLC trace (similar to **b)**)

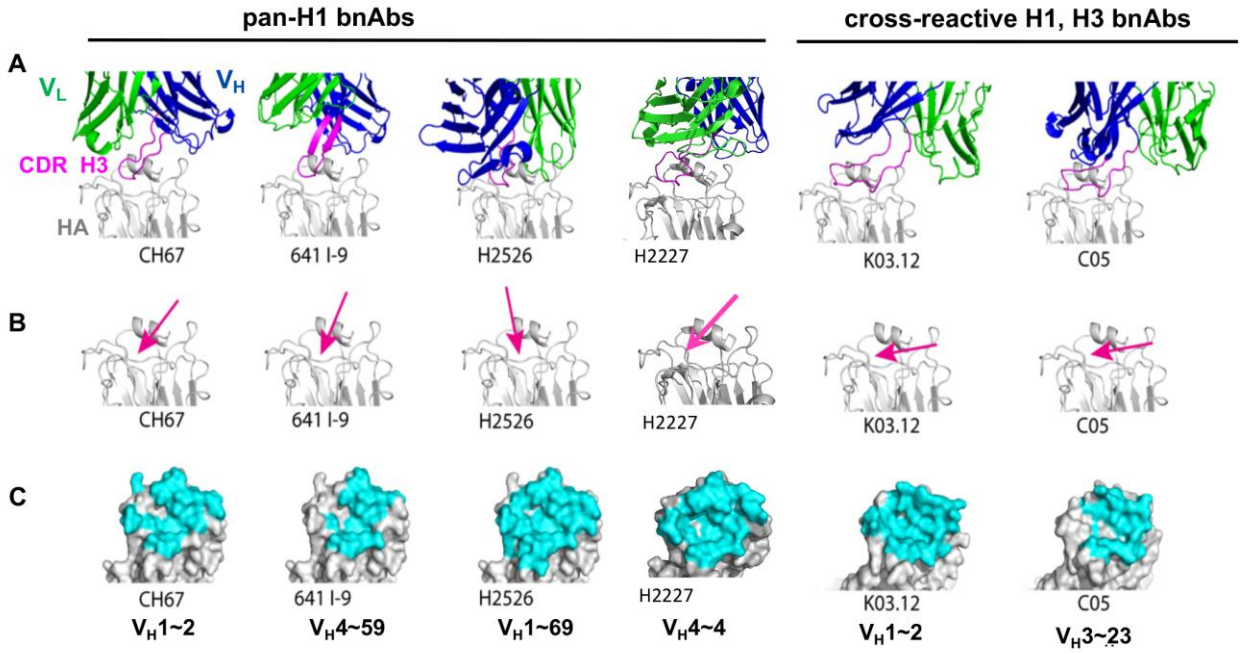


Figure S4: Antibody footprints of RBS-directed antibodies. RBS-directed antibodies used in this study to obtain ELISA and BLI binding affinities. **A,** Crystal structures of CH67 (PDB 4HKX), 641 I-9 (PDB 4YK4), H2526 (PDB 4YJZ), H2227 (PDB 6Q1K), K03.12 Fab (PDB 5W08) and C05 (PDB 4FQR) in complex with HA (silver). The variable heavy and light domains are colored blue and green, respectively with the CDR H3 in the antigen combining site shown in magenta. **B,** An approximate angle of approach of the CDR H3 of each antibody with the HA RBS. **C,** The overall footprint of each antibody with HA is shown in cyan with variable heavy (V_H) gene usage noted. All figures were created using PyMOL.

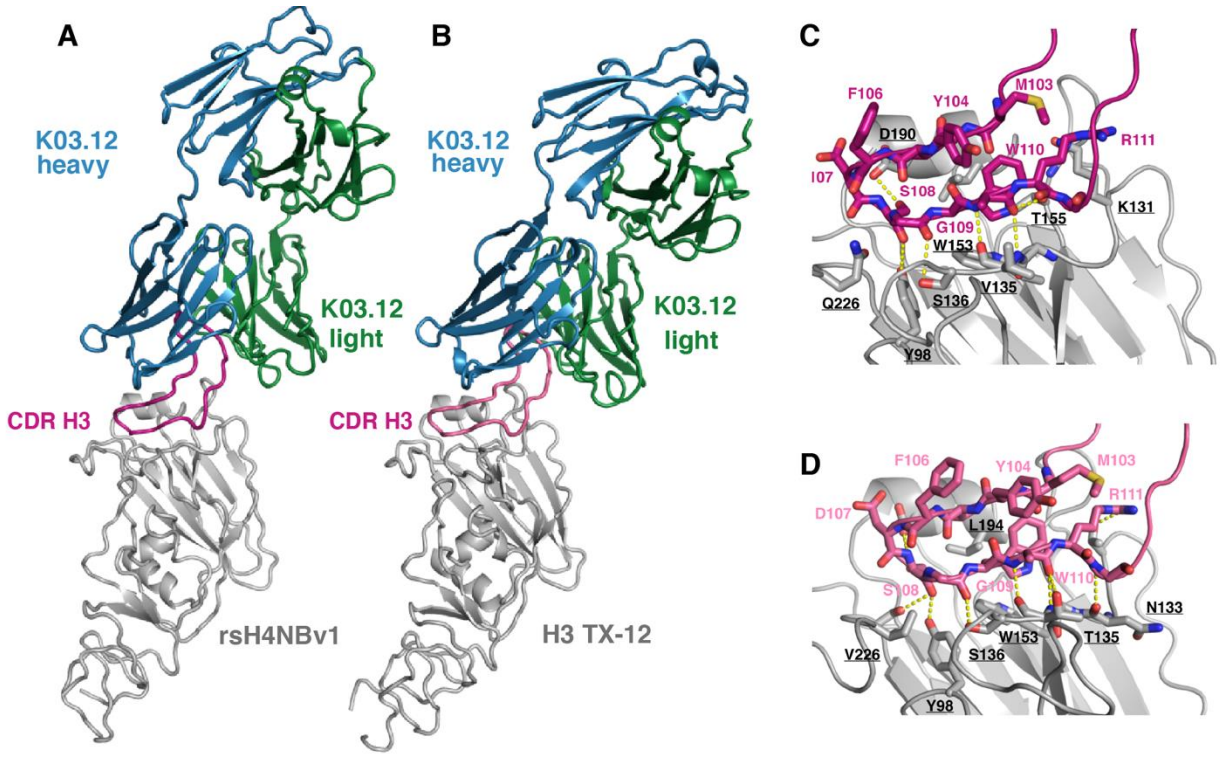


Figure S5: A, K03.12 Fab (heavy and light chains are colored blue and green, respectively) in complex with rsH4NBv1 HA1 "head" (silver). The CDR H3 (magenta) is marked. **B,** K03.12 Fab (heavy and light chains are colored blue and green, respectively) in complex with H3 TX-12 (PDB 5W08) HA1 "head" (silver). The CDR H3 (magenta) is marked. **C, D,** Close-up of the antigen binding site for rsH4NBv1 (c) and H3 TX-12 (d). The CDR H3 (magenta) is shown in sticks with key interacting HA residues (silver). Hydrogen bonds are denoted in yellow, dashed-lines.

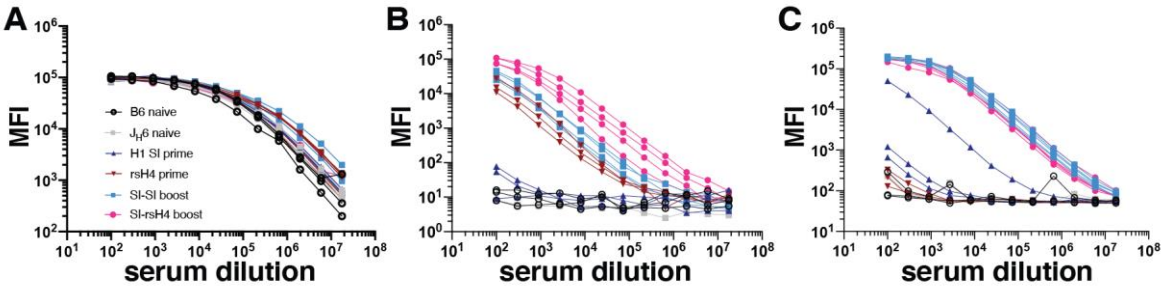


Figure S6: A, Total IgG titer measurements of the different mouse cohorts. The sera from the different groups were serially diluted and the binding to the coated HAs measured by the Luminex assay as described in Methods. **B, C,** Serum reactivity measurements to rsH4NBv3 and H1 SI-06 HAs, respectively, using Luminex assay.

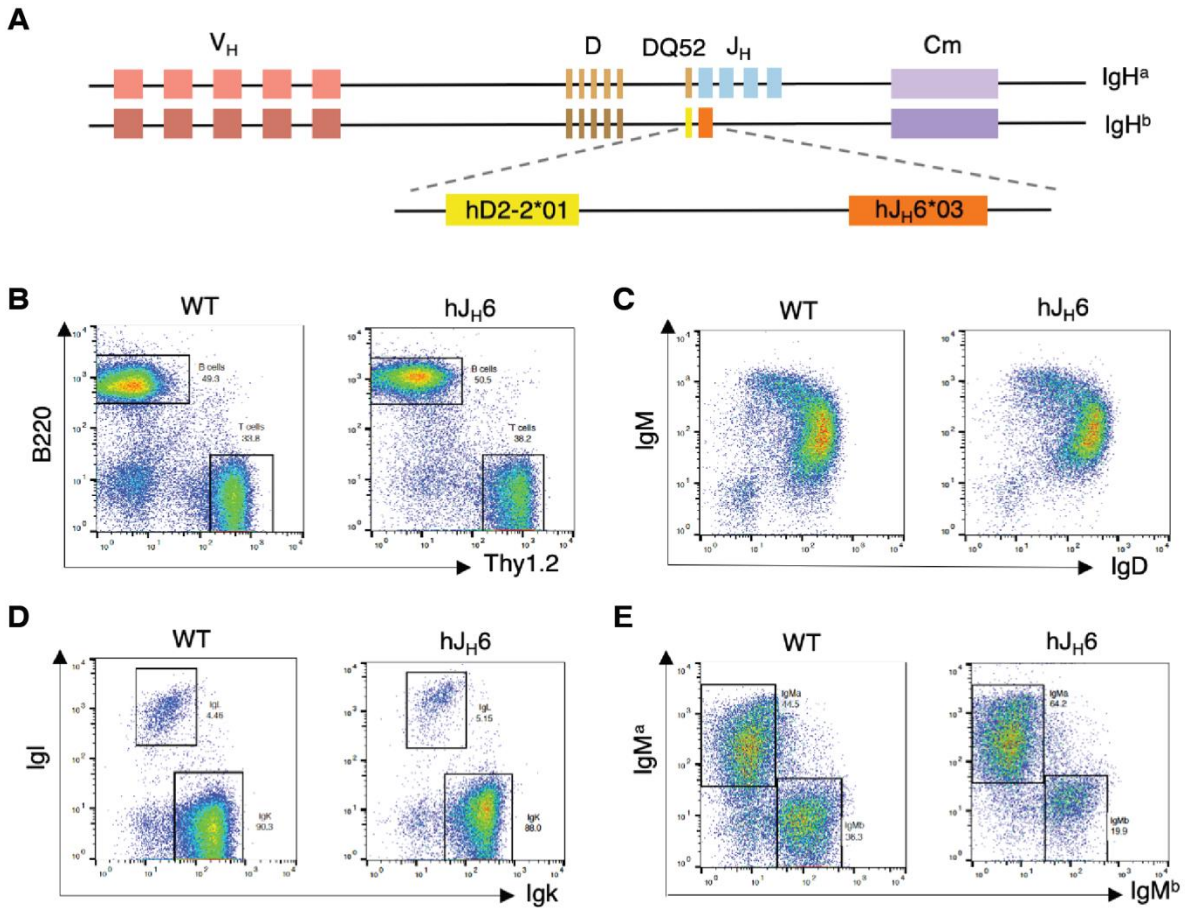


Figure S7: A, Diagram to illustrate the integration of human D2-2*01 (hD2-2*01) and human J_H6*03 (hJ_H6*03) segments into the mouse IgH locus. The diagram is for illustration purpose only and is not drawn to scale. **B,** FACS analysis of B cells and T cells in a wild-type (WT) control mouse and a human JH6 (hJ_H6) mouse. The plots are gated on the lymphocyte population in the FSC/SSC plot. The surface markers detected in this FACS plot are shown. The percentages of B and T cells are shown next to the gates. **C,** FACS analysis of IgM and IgD expression on splenic B cells from WT or hJ_H6 mice. The plots are gated on $B220^+$ B cells. **D,** FACS analysis of IgM and IgD expression on splenic B cells from WT or hJ_H6 mice. The plots are gated on $B220^+$ B cells. **e,** FACS analysis of IgM^a and IgM^b expression on splenic B cells from WT or hJ_H6 mice. Because the WT mouse was of mixed 129/Sv and C57BL/6 genetic background, the mouse had similar proportions of IgM^a and IgM^b B cells. The under-representation of IgM^b B cells in hJ_H6 mouse was presumably due to the presence of only one hJ_H6 segment on this IgH allele. The plots are gated on $B220^+$ B cells.

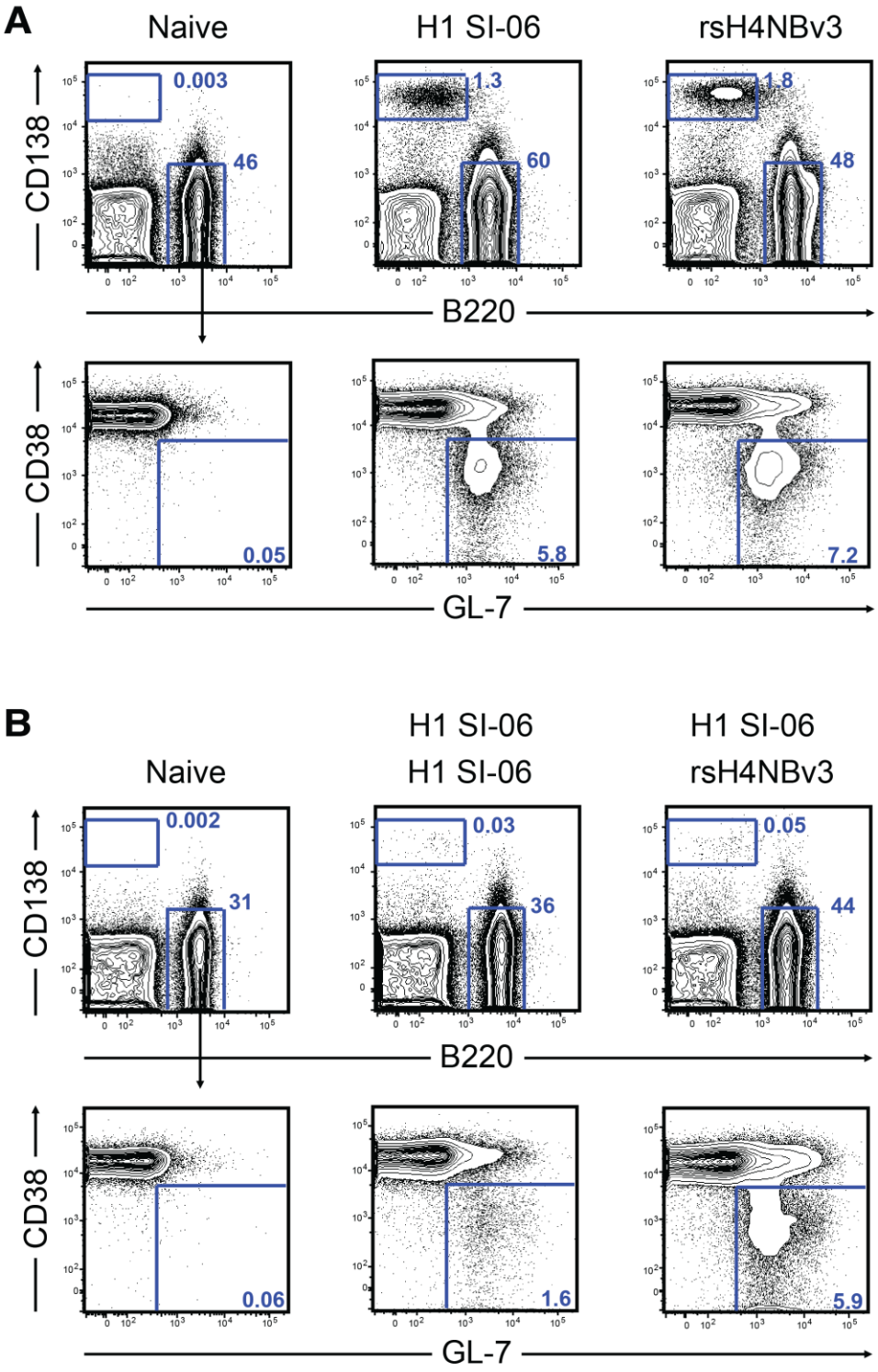
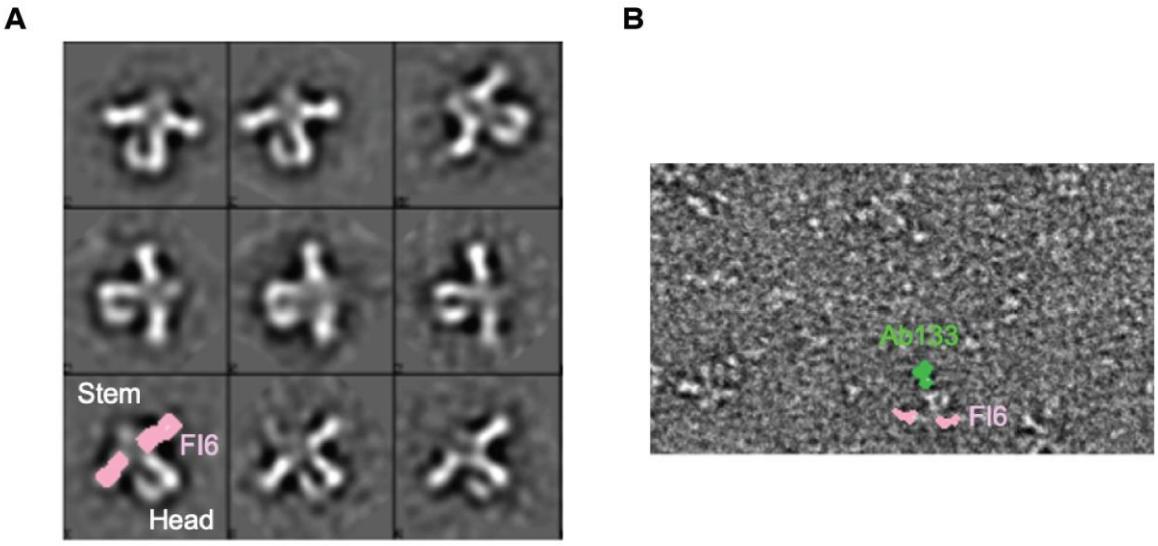


Figure S8: A, Mice were immunized with recombinant H1 SI-06 or rsH4NBv3 HA in Alhydrogel® via footpad and the GC responses were assessed by flow cytometry 8 days after immunization. **B,** after initial immunization as in **A**, the mice were boosted with H1 SI-06 or rsH4NBv3 56 days later; the GC responses were assessed by flow cytometry 8 days after boost. Representative flow diagrams of B cells in the popliteal lymph nodes are shown.



C

```
Ab144_HC      EVQLQESGSELVKPGASVKISCKASGYEFSSSWMNWVKQRPGKGLEWIGRISPGDGDTNY
Ab133_HC      EVQLQESGPDLVKPGASVKISCKASGYEFSSSWMNWVKQRPGKGLEWIGRISPGDGDTNY
              ******** :**************************************************

Ab144_HC      NGEFKGKATLTADKSSTTAYMQLTSLTSEDSAVYFCARSDQFYFYYFYMDVWGKGTTVTV
Ab133_HC      NGEFKGKATLTADTSSTTAYMQLTSLTSEDSAVYFCARSDHYYFYYFYMDVWGKGTTVTV
              *************.***************************::******************


Ab144_LC      DIVMTQTPLSLPVSLGDQASISCRSSQPIVHSSGNTYLEWYLQKPGQSPKLLIYKVSNRF
Ab133_LC      DIVMTQTPLSLPVSLGDQASISCRSSQPIVHSSGNTYLEWYLQKPGQSPKLLVYKVSNRF
              ************************************************:*******

Ab144_LC      SGVPDRFSGSGSGTDFTLKISRVEAEDLGIYYCFQASHVPWTFGGGTKLEIK
Ab133_LC      SGVPDRFSGSGSGTDFTLKISRVEAEDLGIYYCFQGSHVPWTFGGGTKLEIK
              ********************************.***************
```

Figure S9. Head-binding by Ab133. **A**, 2D class averages from negative stain EM are shown for HA in complex with the stem antibody, FI6. The Fab is colored in pink. **B**, Portion of a negative stain EM micrograph is shown. Ab133 is colored green and FI6 is colored in pink. Complexes visualized are in the absence of crosslinking agents. **C**, Sequence alignment of Ab133 and Ab144. The variable heavy and light chains are shown. Legend: (*) fully conserved residue, (:) strong conservation of properties, (.) weak conservation of properties, a blank space indicates little to no conservation of properties.

Table S1. Crystallographic Data Collection and Model Refinement statistics.
Statistics for the highest-resolution shell are shown in parentheses.

Data collection and processing	
Wavelength (Å)	0.999
Resolution range (Å)	46.05 - 4.0
Space group	C 1 2 1
Unit cell a,b,c (Å) α,β,γ (°)	92.3, 141.5, 166.9 90, 102.6, 90
Total reflections	63073 (4479)
Unique reflections	17320 (1493)
Multiplicity	3.6 (3.0)
Completeness (%)	97.5 (85.8)
Mean I/sigma(I)	4.51 (1.52)
R_{merge}	0.29 (0.73)
R_{meas}	0.34 (0.89)
R_{pim}	0.17 (0.49)
CC1/2	0.95 (0.54)
CC*	0.99 (0.84)
Refinement	
R_{work}/R_{free}	0.29/0.32
RMS bonds (Å)/angles(°)	0.003/0.70
Ramachandran favored/outliers (%)	94.65/0.14
Average B-factor	97

Table S2. V_H, D, J_H usage of IgM^b B cells.

This table lists examples of the composition of V_H, D, J_H segments in HCs expressed by sorted IgM^b splenic B cells. hD2-2 and hJH6 are highlighted with yellow and red. "-" represents D remnants that are too short to define the germline segment.

VH	D	JH	CDR H3 (aa)
VH1-55	hD2-2	hJH6	24
VH1-22	hD2-2	hJH6	23
VH1-82	hD2-2	hJH6	22
VH3-1	hD2-2	hJH6	22
VH8-8	hD2-2	hJH6	22
VH1-18	hD2-2	hJH6	20
VH1-9	mD2-9	hJH6	20
VH1-81	hD2-2	hJH6	20
VH1-75	mD2-9	hJH6	20
VH5-17	mD3-2	hJH6	18
VH1-55	mD1-1	hJH6	17
VH3-6	mD2-1	hJH6	17
VH1-26	mD1-1	hJH6	17
VH1-39	hD2-2	hJH6	16
VH1-75	hD2-2	hJH6	16
VH1-63	mD2-1	hJH6	14
VH1-72	mD2-5	hJH6	13
VH1-55	mD2-5	hJH6	13
VH2-9	-	hJH6	13
VH1-64	mD2-4	hJH6	13
VH1-58	mD2-5	hJH6	13
VH9-3	-	hJH6	11
VH1-42	-	hJH6	11
VH1-76	mD2-5	hJH6	11
% hD2-2, JH6	38% hD2-2	100% hJH6	

Table S3. Protein sequences of the different HA constructs.

>H4 A/duck/Czechoslovakia/1956(GenBank: AAA43216.1) EISSQNYTGNPVICLGHHAVPNGTMVKTLTDDQIEVVAAQELVESQHLPELCPSPLRLVDGQTCDIVNGA LGSPGCDHLNGAEWDIFIERPTAVDTCYPFDVPDYQSLRSILANNGKFEFIAEEFQWSTVKQNGKSGACK RANVNDFFNRLNWLTKSDGNAYPLQNLTKINNGDYARLYIWGVHHPSTDTEQTNLYKNNPGRVTVSTK TSQTSVVPNIGSRPWVRGQSGRISFYWTIVEPGDLIVFNTIGNLIAPRGHYKLNSQKKSTILNTAVPIGSCV SKCHTDRGSITTTKPFQNISRISIGDCPKYVKQGSLKLATGMRNIPEKATRGLFGAIAGFIENGWQGLIDG WYGFRHQNAEGTGTAADLKSTQAAIDQINGKLNRLIEKTNEKYHQIEKEFEQVEGRIQDLEKYVEDTKI DLWSYNAELLVALENQHTIDVTDSEMNKLFERVRRQLRENAEDKGNGCFEIFHQCDNNCIESIRNGTYD HDIYRDEAINNRFQIQGVKLTQG
>rsH4v1 EISSQNYTGNPVICLGHHAVPNGTMVKTLTDDQIEVVAAQELVESQHLPELCPSPLRLVDGQTCDIVNGA LGSPGCDHLNGAEWDIFIERPTAVDTCYPFDVPDYQSLRSILANNGKFEFIAEEFQWSTVKQNGVSASCS RANVNDFFNRLNWLTGKNGLYPLQNLTKINNGDYARLYIWGVHHPSNIGDQTNLYKNNPGRVTVSTKT SQTSVVPNIGSRPKVRDQEGRISFYWTIVEPGDLIVFNTIGNLIAPRGHYKLNSQKKSTILNTAVPIGSCVS KCHTDRGSITTTKPFQNISRISIGDCPKYVKQGSLKLATGMRNIPEKATRGLFGAIAGFIENGWQGLIDGW YGFRHQNAEGTGTAADLKSTQAAIDQINGKLNRLIEKTNEKYHQIEKEFEQVEGRIQDLEKYVEDTKIDL WSYNAELLVALENQHTIDVTDSEMNKLFERVRRQLRENAEDKGNGCFEIFHQCDNNCIESIRNGTYDHD IYRDEAINNRFQIQGVKLTQG
>rsH4v2 EISSQNYTGNPVICLGHHAVPNGTMVKTLTDDQIEVVAAQELVESQHLPELCPSPLRLVDGQTCDIVNGA LGSPGCDHLNGAEWDIFIERPTAVDTCYPFDVPDYQSLRSILANNGKFEFIAEEFQWSTVTQNGVSASCS RANVNDFFNRLNWLTGKNGLYPLQNLTKINNGDYARLYIWGVHHPSNIGDQRALYKNNPGRVTVSTKT SQTSVVPNIGSRPKVRDQEGRISFYWTIVEPGDLIVFNTIGNLIAPRGHYKLNSQKKSTILNTAVPIGSCVS KCHTDRGSITTTKPFQNISRISIGDCPKYVKQGSLKLATGMRNIPEKATRGLFGAIAGFIENGWQGLIDGW YGFRHQNAEGTGTAADLKSTQAAIDQINGKLNRLIEKTNEKYHQIEKEFEQVEGRIQDLEKYVEDTKIDL WSYNAELLVALENQHTIDVTDSEMNKLFERVRRQLRENAEDKGNGCFEIFHQCDNNCIESIRNGTYDHD IYRDEAINNRFQIQGVKLTQG
>rsH4v3 EISSQNYTGNPVICLGHHAVPNGTMVKTLTDDQIEVVAAQELVESQHLPELCPSPLRLVDGQTCDIVNGA LGSPGCDHLNGAEWDIFIERPTAVDTCYPFDVPDYQSLRSILANNGKFEFIAEEFQWSTVTQNGVSASCS RANVSDFFNRLNWLTGKNGLYPLQNLTKINNGDYARLYIWGVHHPPNIGDQRALYHNEPGRVTVSTKT SQTSVVPNIGKRPKVRDQEGRISFYWTIVEPGDLIVFNTIGNLIAPRGHYKLNSQKKSTILNTAVPIGSCVS KCHTDRGSITTTKPFQNISRISIGDCPKYVKQGSLKLATGMRNIPEKATRGLFGAIAGFIENGWQGLIDGW YGFRHQNAEGTGTAADLKSTQAAIDQINGKLNRLIEKTNEKYHQIEKEFEQVEGRIQDLEKYVEDTKIDL WSYNAELLVALENQHTIDVTDSEMNKLFERVRRQLRENAEDKGNGCFEIFHQCDNNCIESIRNGTYDHD IYRDEAINNRFQIQGVKLTQG
>H14 A/mallard/Wisconsin/10OS3941/2010_(GenBank AGE03043) QITNGNTGNPVICLGHHAVENGTSVKTLTDNHIEVVSAKELVETNHINELCPSPLKLVDGQDCDLINGAL GSPGCDHLQDTTWDVFIERPTAMDTCYPFDVPDYQSLRSILASSGSLEFIAEQFTWNGVTVDGSSSACLR GGRNGFFTRLNWLTRVKNGNYGPINVTKENTGSYVRLYLWGVHHPSSDTEQTDLYKVATGRVTVSTRS DQISIIPNIGSRPRVRNQSGRISIYWTLVNPGDSIIFNSIGNLIAPRGHYKINKSTKGTVLKSDKKIGSCTSPC LTDKGSIQSDKPFQNVSRIAIGNCPKYVKQGSLMLATGMRNIPDKQTKGLFGAIAGFIENGWQGLIDGW YGFRHQNAEGTGTAADLKSTQAAIDQINGKLNRLIEKTNEKYHQIEKEFEQVEGRIQDLEKYVEDTKIDL WSYNAELLVALENQHTIDVTDSEMNKLFERVRRQLRENAEDQGNGCFEIFHQCDNNCIESIRNGTYDHN IYRDEAINNRIKINPVNLTMG
>rsH14v1 QITNGNTGNPVICLGHHAVENGTSVKTLTDNHIEVVSAKELVETNHINELCPSPLKLVDGQDCDLINGAL GSPGCDHLQDTTWDVFIERPTAMDTCYPFDVPDYQSLRSILASSGSLEFIAEQFTWNGVTVDGVSASCSR GGRNGFFTRLNWLTGKNGLYGPINVTKENTGSYVRLYLWGVHHPSNIGDQTSLYKVATGRVTVSTRSD QISIIPNIGSRPRVRDQEGRISIYWTLVNPGDSIIFNSIGNLIAPRGHYKINKSTKGTVLKSDKKIGSCTSPCL TDKGSIQSDKPFQNVSRIAIGNCPKYVKQGSLMLATGMRNIPDKQTKGLFGAIAGFIENGWQGLIDGWY GFRHQNAEGTGTAADLKSTQAAIDQINGKLNRLIEKTNEKYHQIEKEFEQVEGRIQDLEKYVEDTKIDLW SYNAELLVALENQHTIDVTDSEMNKLFERVRRQLRENAEDQGNGCFEIFHQCDNNCIESIRNGTYDHNIY RDEAINNRIKINPVNLTMG

>rsH14v2
QITNGNTGNPVICLGHHAVENGTSVKTLTDNHIEVVSAKELVETNHINELCPSPLKLVDGQDCDLINGAL
GSPGCDHLQDTTWDVFIERPTAMDTCYPFDVPDYQSLRSILASSGSLEFIAEQFTWNGVTVDGVSASCSR
GGRSGFFTRLNWLTGKNGLYGPINVTKENTGSYVRLYLWGVHHPSNIGDQRALYHVETGRVTVSTRSD
QISIIPNIGKRPRVRDQEGRISIYWTLVNPGDSIIFNSIGNLIAPRGHYKINKSTKGTVLKSDKKIGSCTSPCL
TDKGSIQSDKPFQNVSRIAIGNCPKYVKQGSLMLATGMRNIPDKQTKGLFGAIAGFIENGWQGLIDGWY
GFRHQNAEGTGTAADLKSTQAAIDQINGKLNRLIEKTNEKYHQIEKEFEQVEGRIQDLEKYVEDTKIDLW
SYNAELLVALENQHTIDVTDSEMNKLFERVRRQLRENAEDQGNGCFEIFHQCDNNCIESIRNGTYDHNIY
RDEAINNRIKINPVNLTMG