Polypeptides for treating and/or limiting influenza infection

Cross Reference

This application claims priority to U.S. Provisional Application Serial Nos. 61/370,410 filed August 3, 2010; 61/436,058 filed January 25, 2011; 61/440,771 filed February 8, 2011; and 61/485,395 filed May 12, 2011, each of which is incorporated herein by reference in its entirety.

Statement of Government Support

This invention was made with government support under grant number 5P41RR011823-15 awarded by National Institutes of Health and grant number HR0011-08-0085 awarded by Defense Advanced Research Projects Agency and grant number HDTRA1-10-1-0040 awarded by Defense Threat Reduction Agency. The government has certain rights in the invention.

Background

Influenza virus is a member of Orthomyxoviridae family. There are three subtypes of influenza viruses designated A, B, and C. The influenza virion contains a segmented negative-sense RNA genome, encoding, among other proteins, hemagglutinin (HA) and neuraminidase (NA). Influenza virus infection is initiated by the attachment of the virion surface HA protein to a sialic acid-containing cellular receptor (glycoproteins and glycolipids). The NA protein mediates processing of the sialic acid receptor, and virus penetration into the cell depends on HA-dependent receptor-mediated endocytosis. In the acidic confines of internalized endosomes containing an influenza virion, the HA2 protein undergoes conformational changes that lead to fusion of viral and cell membranes and virus uncoating and M2-mediated release of M1 proteins from nucleocapsid-associated ribonucleoproteins (RNPs), which migrate into the cell nucleus for viral RNA synthesis. Antibodies to HA proteins prevent virus infection by neutralizing virus infectivity.

Influenza presents a serious public-health challenge and new therapies are needed to combat viruses that are resistant to existing antivirals or escape neutralization by the immune system.

Summary of the Invention

In a first aspect, the present invention provides polypeptides comprising an amino acid sequence according to general formula I

R1-R2-Phe-R3-R4-R5-R6-R7-R8-R9-R10-R11-R12-R13-R14-R15-R16 (SEQ ID NO: 1), wherein

R1 is selected from the group consisting of Ser, Ala, Phe, His, Lys, Met, Asn, Gln, Thr, Val, Tyr, and Asp;

R2 can be any amino acid;

R3 is selected from the group consisting of Asp, Ala, Glu, Gly, Asn, Pro, Ser, and Tyr;

R4 is selected from the group consisting of Leu and Phe;

R5 can be any amino acid;

R6 is selected from the group consisting of Met, Phe, His, Ile, Leu, Gln, and Thr;

R7 is selected from the group consisting of Arg, Gly, Lys, Gln, and Thr;

R8 is selected from the group consisting of Ile, Asn, Gln, Val, and Trp;

R9 is selected from the group consisting of Met, Gly, Ile, Lys, Leu, Asn, Arg, Ser, Thr, Val, His, and Tyr;

R10 is selected from the group consisting of Trp and Phe;

R11 is selected from the group consisting of Ile, Phe, Ser, Thr, and Val;

R12 is selected from the group consisting of Tyr, Cys, Asp, Phe, His, Asn, and Ser;

R13 is selected from the group consisting of Val, Ala, Phe, Ile, Leu, Asn, Gln, Thr, and Tyr;

R14 is selected from the group consisting of Phe, Glu, and Leu;

R15 is selected from the group consisting of Ala, Gly, Lys, Arg, and Ser; and

R16 is selected from the group consisting of Phe, Cys, His, Lys, Leu, Met, Asn, Gln, Arg, Thr, Val, Trp, and Tyr.

In one embodiment, the polypeptide comprises or consists of R1-R2-Phe-R3-R4-R5-R6-R7-R8-R9-R10-R11-R12-R13-R14-R15-R16-X1-R17 (SEQ ID NO: 2), wherein

X1 is 4-8 amino acids in length, wherein each position can be any amino acid; and

R17 is Phe or Tyr.

In another aspect, the present invention provides polypeptides comprising an amino acid sequence according to general formula II

R1-R2-R3-R4-R5-R6-R7-R8-R9-Ala-R10-R11-Phe (SEQ ID NO: 83), wherein

R1 is selected from the group consisting of Phe and Val;

R2 is selected from the group consisting of Ser, Ala, Phe, Gly, Ile, Lys, Leu, Met, Asn, Pro, Gln, Arg, Thr, and Val;

R3 is selected from the group consisting of Glu, and Asp;

R4 is selected from the group consisting of Asn, His, Ile, Lys, Leu, Met, Arg, Ser, and Thr;

R5 is selected from the group consisting of Leu, Phe, Ile, Met, Asn, Gln, and Val;

R6 is selected from the group consisting of Ala, Asp, Lys, Met, Asn, Gln, Arg, Glu, and Val;

R7 is selected from the group consisting of Phe, Asp, Asn, and Tyr;

R8 is selected from the group consisting of Glu, Ala, Asp, Gly, His, Lys, Leu, Met, Asn, Gln, Arg, Ser, Thr, Val, and Trp;

R9 is selected from the group consisting of Leu, Phe, Ile, Met, and Val;

R10 is selected from the group consisting of Leu, Ile, Met, and Tyr; and

R11 is selected from the group consisting of Ser, Ala, Gly, and Tyr;

In one embodiment, the polypeptides of general formula II comprise or consist of R1-R2-R3-R4-R5-R6-R7-R8-R9-Ala-R10-R11-Phe-**X1-R12-R13-X2-R14 (SEQ ID NO: 84)**, wherein

X1 is 5-15 amino acids in length, wherein each position can be any amino acid;

R12 is selected from the group consisting of Gln, Tyr, Phe, Met, Arg, Lys, and Gly;

R13 is selected from the group consisting of Tyr, Asp, Met, Asn, and Ser;

X2 is any amino acid; and

R14 is selected from the group consisting of Ser, Arg, and Lys.

In another aspect, the present invention provides polypeptides comprising an amino acid sequence selected from the group consisting of

(a) MADTLLILGDSLSAGYQMLAEFAWPFLLNKKWSKTSVVNASISGDTSQQGL ARLPALLKQHQPRWVLVELGGNDGLEGFQPQQTEQTLRQILQDVKAANAEP LLMQIRPPANYGRRYNEAFSAIYPKLAKEFDVPLLPFFMEEVYLKPQWMQDD GIHPNYEAQPFIADWMAKQLQPLVNH (SEQ ID NO: 155);

(b) MAETKNFTDLVEATKWGNSLIKSAKYSSKDKMAIYNYTKNSSPINTPLRSAN GDVNKLSENIQEQVRQLDSTISKSVTPDSVYVYRLLNLDYLSSITGFTREDLH MLQQTNEGQYNSKLVLWLDFLMSNRIYRENGYSSTQLVSGAALAGRPIELKL ELPKGTKAAYIDSKELTAYPGQQEVLLPRGTEYAVGTVELSKSSQKIIITAVVF KK (SEQ ID NO: 140); and

(c) MFTGVIIKQGCLLKQGHTRKNWSVRKFILREDPAYLHYYYPLGYFSPLGAIHL RGCVVTSVESEENLFEIITADEVHYFLQAATPKERTEWIKAIQMASR (SEQ ID NO: 211).

In a third aspect, the present invention provides isolated nucleic acids encoding the polypeptide of any embodiment of the invention. In a fourth aspect, the present invention provides recombinant expression vectors comprising the nucleic acid of the third aspect of the invention, operatively linked to a suitable control sequence. In a fifth aspect, the present invention provides recombinant host cells comprising the recombinant expression vectors of the fourth aspect of the invention. In a sixth aspect, the present invention provides antibodies that selectively bind to the polypeptides of the invention.

In a seventh aspect, the present invention provides pharmaceutical compositions, comprising one or more polypeptides according of the invention and a pharmaceutically acceptable carrier.

In an eighth aspect, the present invention provides methods for treating and/or limiting an influenza infection, comprising administering to a subject in need thereof a therapeutically effective amount of one or more polypeptides of the invention, salts thereof, conjugates thereof, or pharmaceutical compositions thereof, to treat and/or limit the influenza infection.

In a ninth aspect, the present invention provides methods for diagnosing an influenza infection, or monitoring progression of an influenza infection, comprising

(a) contacting a biological sample from a subject suspected of having an influenza infection with a diagnostically effective amount of one or more polypeptides of the invention under conditions suitable for binding of the polypeptide to a viral HA protein present in the sample; and

(b) detecting polypeptide-viral HA binding complexes,

where the presence of such binding complexes indicates that the subject has an influenza infection, or provides a measure progression of an influenza infection.

In a tenth aspect, the present invention provides methods for identifying candidate influenza vaccines, comprising

(a) contacting test compounds with a polypeptide of the present invention under conditions suitable for polypeptide binding;

(b) removing unbound test compounds; and

(c) identifying those test compounds that bind to the polypeptide of the invention, wherein such test compounds are candidate influenza vaccines.

In an eleventh aspect, the present invention provides methods for identifying candidate compounds for treating, limiting, and/or diagnosing influenza infection, comprising

(a) contacting an influenza HA protein with (i) test compounds and (ii) a polypeptide of the present invention, under conditions suitable for binding of the HA protein to the polypeptide of the present invention; and

(b) identifying those test compounds that outcompete the polypeptide for binding to the HA protein, wherein such test compounds are candidate compounds for treating, limiting, and/or diagnosing influenza infection.

Description of the Figures

Fig. 1. Overview of the design process. The flow chart illustrates key steps in the design process for novel binding proteins, with thumbnails illustrating each step in the creation of binders that target the stem of the 1918 HA.

Fig. 2. Design of HB36 and HB80, targeting the stem of the 1918 HA. (**A**) Surface representation of the trimeric HA structure (PDB 3R2X) from the 1918 pandemic virus. Broadly neutralizing antibody CR6261 binds a highly conserved epitope in the stem region, close to the viral membrane (bottom). (**B**) Enlarged view of the CR6261 epitope , with CR6261 contact

residues depicted as sticks. This target site on HA contains a groove lined by multiple hydrophobic residues. Loops on either side of this hydrophobic groove (above and below) constrain access to this region. Key residues on HA2 are noted in one-letter code. (**C and D**) Front view of the designed interaction between HB36 (C) and HB80 (D) and the target site on HA. HA is rotated approximately 60° relative to Fig. 2A. HB36 and HB80 residues are depicted as sticks, with hotspot residues noted (F49 and M53 for HB36 and L21, F25, and Y40 for HB80). For clarity, the non-contacting regions from the designs have been omitted. (**E and F**) Further details of the designed interactions of HB36 (E) and HB80 (F) with 1918/H1 HA. (**G and H**) Initial binding data for HB36 (G) and HB80 (H) designs (before affinity maturation). When incubated with 1 uM 1918 HA, yeast displaying the two designed proteins show an increase in fluorescent phycoerythrin signal (x-axis) compared to the absence of 1918 HA.

Fig. 3 Affinity maturation. Substitutions that increase the affinity of the original designs can be classified as deficiencies in modeling the **(A and B)** repulsive interactions HB36 Ala60Val (A), HB80 Met26Thr (B); **(C and D)** electrostatics HB36 Asn64Lys (C), HB80 Asn36Lys (D); **(E and F)** and solvation HB36 Asp47Ser (E), HB80 Asp12Gly (F). Binding titrations of HB36.4 (**G**) and HB80.3 (**H**) to SC1918/H1 HA as measured by yeast surface display. Circles represent the affinity-matured design, Squares the scaffold protein from which the design is derived, and crosses represent the design in the presence of 750nM inhibitory CR6261 Fab.

Fig. 4 Crystal structure of HB36.3-SC1918/H1 complex validates the precision of the computational design. (**A**) Superposition of the crystal structure of HB36.3-SC1918/H1 complex and the computational design reveals good agreement in the position of the main recognition helix, with a slight rotation of the rest of the protein domain. Superposition was performed using the HA2 subunits. For clarity, only the HA from the crystal structure is depicted here (the HA used for superposition of the design, which is essentially identical to the crystal structure, was omitted). (**B**) Close up of the SC1918 HA-HB36.3 interface, highlighting the close agreement between the design and the crystal structure. The main recognition helix is oriented approximately as in (A). (**C**) Unbiased 2Fo-Fc (**gray** mesh, contoured at 1σ) and Fo-Fc (dark mesh, contoured at 3σ) electron-density maps for the main recognition helix of HB36.3. The helix is oriented as in (B), with key contact residues of the left face of the helix in this view labeled (the right surface faces and interacts with the core of the HB36.3 protein). Significant density was observed for most of the large side chains at the interface with HA, including F49,

M53, W57, F61, and F69 (not visible in this view). While side chains are shown here to illustrate their agreement with the experimental electron density, maps were calculated after initial refinement of an HA-HB36.3 model with the following side chains truncated to alanine (no prior refinement with side chains present): F49, M53, M56, W57, F61, and F69.

Fig. 5. HB80.3 binds and inhibits multiple HA subtypes. **(A)** Phylogenetic tree depicting the relationship between the 16 influenza A hemagglutinin subtypes. These subtypes can be divided into two main lineages, groups 1 and 2. CR6261 has broad activity against group 1 viruses. HB80.3 has a similar cross-reactivity profile and binds multiple group 1 subtypes, including H1 and H5. **(B)** Binding data for HB80.3 and CR6261 Fab against a panel of HAs. "+", "++", and "+++" indicate relative degree of binding (approximately 10^{-7}, 10^{-8}, and 10^{-9} M, respectively), while "-" indicates no detectable binding at the highest concentration tested (100nM). **(C)** HB80.3 inhibits the pH-induced conformational changes that drive membrane fusion. Exposure to low pH converts 1918 H1 HA (top panel) and the Viet04 H5 HA to a protease susceptible state (lane 1), while HAs maintained at neutral pH are highly resistant to trypsin (lane 3). Pre-incubation of HB80.3 with H1 and H5 prevents pH-induced conformational changes and retains the HAs in the protease-resistant, pre-fusion state (lane 2).

Figure 6. Binding titrations of HB36 to SC1918/H1 HA as measured by yeast surface display. Circles represent the computational design, squares the scaffold protein from which the design is derived, and crosses represent the design in the presence of 1.5uM inhibitory CR6261 Fab.

Figure 7. Phycoerythrin (PE) intensity histograms for (a.) HB80 design and (b.) the scaffold the design was derived from (PDB code 2CJJ). Dashed lines represent the population of yeast cells displaying the design in the absence and dark lines the presence of 1uM H1 HA.

Figure 8. Truncation after position 54 on HB80 M26T N36K increases mean surface display. FITC intensity histograms of (a.) HB80 M26T N36K and (b.) HB80 M26T N36K Δ54-95. In both cases, gray lines represent unlabeled cells, while black lines represent cells labeled with anti-cmyc FITC.

Figure 9. Alanine scanning mutagenesis of key residues at the designed interface of HB36.3 (a.) and HB80.1 (b.) completely abrogating binding. Binding was measured by yeast surface display titrations

Figure 10. Yeast display titrations of designs to H1 & H5 HA subtypes show heterosubtypic binding of (A.) HB36.4 & (B.) HB80.3 design variants. For both panels, circles are binding titrations of SC/1918/H1 HA and squares the titration data for VN/2004/H5 HA.

Figure 11. Protease susceptibility-inhibition assay for HB36.3 against the SC1918/H1 HA. **(A)** The upper panel shows the effect of various pH treatments and trypsin digestion on SC1918 HA alone. Most of the HA is converted to the protease-susceptible, post-fusion conformation below pH ~6.0-6.5. The lower panel shows the identical assay for the HB36.3-SC1918 complex (saturated with HB36.3 and purified by gel filtration prior to the experiment; approximately 1:1 molar ratio of HB36.3 to HA). Presence of pre-bound HB36.3 in the reactions is unable to block the conversion of HA to the protease-resistant state. **(B)** Assay carried out under conditions identical those used for HB80.3 as presented in Fig. 5C (approximately 10:1 molar ratio of HB36.3 to HA). HB36.3 has no protective effect under these conditions.

Detailed Description of the Invention

All references cited are herein incorporated by reference in their entirety. Within this application, unless otherwise stated, the techniques utilized may be found in any of several well-known references such as: *Molecular Cloning: A Laboratory Manual* (Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press), *Gene Expression Technology* (Methods in Enzymology, Vol. 185, edited by D. Goeddel, 1991. Academic Press, San Diego, CA), "Guide to Protein Purification" in *Methods in Enzymology* (M.P. Deutshcer, ed., (1990) Academic Press, Inc.); *PCR Protocols: A Guide to Methods and Applications* (Innis, et al. 1990. Academic Press, San Diego, CA), *Culture of Animal Cells: A Manual of Basic Technique, 2nd Ed.* (R.I. Freshney. 1987. Liss, Inc. New York, NY), *Gene Transfer and Expression Protocols*, pp. 109-128, ed. E.J. Murray, The Humana Press Inc., Clifton, N.J.), and the Ambion 1998 Catalog (Ambion, Austin, TX).

As used herein, the singular forms "a", "an" and "the" include plural referents unless the context clearly dictates otherwise. "And" as used herein is interchangeably used with "or" unless expressly stated otherwise.

All embodiments of any aspect of the invention can be used in combination, unless the context clearly dictates otherwise.

In a first aspect, the present invention provides polypeptides comprising an amino acid sequence according to general formula I

R1-R2-Phe-R3-R4-R5-R6-R7-R8-R9-R10-R11-R12-R13-R14-R15-R16 **(SEQ ID NO: 1)**, wherein

R1 is selected from the group consisting of Ser, Ala, Phe, His, Lys, Met, Asn, Gln, Thr, Val, Tyr, and Asp;

R2 can be any amino acid;

R3 is selected from the group consisting of Asp, Ala, Glu, Gly, Asn, Pro, Ser, and Tyr;

R4 is selected from the group consisting of Leu and Phe;

R5 can be any amino acid;

R6 is selected from the group consisting of Met, Phe, His, Ile, Leu, Gln, and Thr;

R7 is selected from the group consisting of Arg, Gly, Lys, Gln, and Thr;

R8 is selected from the group consisting of Ile, Asn, Gln, Val, and Trp;

R9 is selected from the group consisting of Met, Gly, Ile, Lys, Leu, Asn, Arg, Ser, Thr, Val, His, and Tyr;

R10 is selected from the group consisting of Trp and Phe;

R11 is selected from the group consisting of Ile, Phe, Ser, Thr, and Val;

R12 is selected from the group consisting of Tyr, Cys, Asp, Phe, His, Asn, and Ser;

R13 is selected from the group consisting of Val, Ala, Phe, Ile, Leu, Asn, Gln, Thr, and Tyr;

R14 is selected from the group consisting of Phe, Glu, and Leu;

R15 is selected from the group consisting of Ala, Gly, Lys, Arg, and Ser; and

R16 is selected from the group consisting of Phe, Cys, His, Lys, Leu, Met, Asn, Gln, Arg, Thr, Val, Trp, and Tyr.

In one embodiment, general formula I is R1-R2-Phe-R3-R4-R5-R6-R7-R8-R9-R10-R11-R12-R13-R14-R15-R16-**X1-R17 (SEQ ID NO: 2)**, wherein R1 through R16 are as defined above, and wherein

X1 is 4-8 amino acids in length, wherein each position can be any amino acid; and

R17 is Phe or Tyr.

In various embodiments, X1 is 4, 5, 6, 7, or 8 amino acids in length. In another embodiment, X1 comprises the amino acid sequence **Z1** -Arg-**Z2**-Ile-Pro **(SEQ ID NO: 3)**, wherein Z1 is Lys or Asn, and Z2 is selected from the group consisting of Lys, Pro, and Thr.

In another embodiment, that can be combined with any other embodiments herein, general formula I is **A1**-R1-R2-Phe-R3-R4-R5-R6-R7-R8-R9-R10-R11-R12-R13-R14-R15-R16-X1-R17-**B1 (SEQ ID NO: 4)**, wherein R1 through R17 and X1 are as defined above, wherein A1 and/or B1 are optionally present, and

wherein A1 comprises the amino acid sequence: MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYD(**D/V/Y**)EA(**A/D**)(**A/K/R**)VL(**Q/K**)AVY (**E/A**)T(**N/D**) **(SEQ ID NO: 5)**; and

B1 comprises the amino acid sequence (**L/A/V**)HA(**Q/P**)KLARRLLELK(**Q/L**)AASSPLP **(SEQ ID NO: 6)**.The inventors have discovered that polypeptides comprising or consisting of the amino acid sequence of general formula I (derived from HB36.4, as described in more detail in the attached) form helices that recognize and are strong binders to Influenza A hemagglutinin ("HA"), such as influenza viruses of phylogenetic group I, preferably influenza A viruses comprising HA of the H1 or H5 subtype. Thus, the polypeptides can be used, for example, to treat and/or limit development of an influenza infection.

In one embodiment, the polypeptide comprises the polypeptide SAFDLAMRIMWIYVFAF (SEQ ID NO:7), SAFDLAMRIMWIYVFAFKRPIPF (SEQ ID NO:8), or a variant including 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more variant positions or SEQ ID NOS. 7 or 8 according to any embodiment of general formula I. In other exemplary embodiments, the polypeptide comprises or consists of a polypeptide selected from the group consisting of (scaffold derived from noted in parentheses):

DAFDLAMRIMWIYVFAFNRPIPF (HB36.2) (SEQ ID NO: 9);

DAFDLAMRIMWIYVFAF (HB36.2) (SEQ ID NO: 10);

SAFDLAMRIMWIYVFAFNRPIPF (HB36.3) (SEQ ID NO: 11);

SAFDLAMRIMWIYVFAF (HB36.3 and HB36.4) (SEQ ID NO: 7);

SAFDLAMRIMWIYVFAFKRPIPF (HB36.4) (SEQ ID NO: 8);

>HB36.4_s4_E03

HAFDLAMRIHWIYVFAF(SEQ ID NO: 15);

HAFDLAMRIHWIYVFAFKRKIPF(SEQ ID NO: 16);

>HB36.4_s4_E05

SAFDLAMRIIWIYVFAY (SEQ ID NO: 17);

SAFDLAMRIIWIYVFAYKRKIPF (SEQ ID NO: 18);

>HB36.4_s4_E06

SAFDLAMRINWIYVFAF (SEQ ID NO: 19);

SAFDLAMRINWIYVFAFKRPIPF (SEQ ID NO: 20);

>HB36.4_s4_E07

SAFDLAMRINWIYVFAF(SEQ ID NO: 21);

SAFDLAMRINWIYVFAFKRKIPF (SEQ ID NO: 22);

>HB36.4_s4_E08

SAFDLAMTIHWIYNFAF (SEQ ID NO: 23);

SAFDLAMTIHWIYNFAFKRKIPF (SEQ ID NO: 24);

>HB36.4_s4_E09

SAFDLAMRINWIYVFAF (SEQ ID NO: 25);

SAFDLAMRINWIYVFAFKRTIPF (SEQ ID NO: 26);

>HB36.4_s4_E10

SAFDLAMRIHWIYIFAF (SEQ ID NO: 27);

SAFDLAMRIHWIYIFAFKRPIPF (SEQ ID NO: 28);

>HB36.4_s4_E11

SAFDLAMRIHWIYNFAF (SEQ ID NO: 29);

SAFDLAMRIHWIYNFAFKRKIPF (SEQ ID NO: 30);

>HB36.4_s4_E12

SAFDLAMRIHWIYNFAY (SEQ ID NO: 31);

SAFDLAMRIHWIYNFAYKRTIPF (SEQ ID NO: 32);

>HB36.4_s4_E13

SAFDLAMRIHWIYNFAF (SEQ ID NO: 33);

SAFDLAMRIHWIYNFAFKRKIPF(SEQ ID NO: 34);

>HB36.4_s4_E14

SAFDLAMRIHWIYIFAF (SEQ ID NO: 35);

SAFDLAMRIHWIYIFAFKRTIPF (SEQ ID NO: 36);

>HB36.4_s4_E17

SAFDLAMRIHWIYNFAF (SEQ ID NO: 37);

SAFDLAMRIHWIYNFAFKRKIPF (SEQ ID NO: 38);

>HB36.4_s4_E18

SAFDLAMKIHWIYNFAF (SEQ ID NO: 39);

SAFDLAMKIHWIYNFAFKRTIPF (SEQ ID NO: 40);

>HB36.4_s4_E19

SAFDLAMKIHWIYIFAF (SEQ ID NO: 41);

SAFDLAMKIHWIYIFAFKRTIPF (SEQ ID NO: 42);

HAFDLAMRIMWIYVFAF (SEQ ID NO: 44);

SAFDLAMKIMWIYVFAF (SEQ ID NO: 45);

SAFDLAMRIHWIYVFAF (SEQ ID NO: 46);

SAFDLAMRINWIYVFAF (SEQ ID NO: 47);

SAFDLAMRIYWIYVFAF (SEQ ID NO: 48);

SAFDLAMRIMWIYFFAF (SEQ ID NO: 49);

SAFDLAMRIMWIYLFAF (SEQ ID NO: 50);

SAFDLAMRIMWIYTFAF (SEQ ID NO: 51);

SAFDLAMRIMWIYNFAF (SEQ ID NO: 52);

SAFDLAMRIMWIYVFAW (SEQ ID NO: 53);

HAFDLAMRIMWIYVFAFKRPIPF (SEQ ID NO: 55);

SAFDLAMKIMWIYVFAFKRPIPF (SEQ ID NO: 56);

SAFDLAMRIHWIYVFAFKRPIPF (SEQ ID NO: 57);

SAFDLAMRINWIYVFAFKRPIPF (SEQ ID NO: 58);

SAFDLAMRIYWIYVFAFKRPIPF (SEQ ID NO: 59);

SAFDLAMRIMWIYFFAFKRPIPF (SEQ ID NO: 60);

SAFDLAMRIMWIYLFAFKRPIPF (SEQ ID NO: 61);

SAFDLAMRIMWIYTFAFKRPIPF (SEQ ID NO: 62);

SAFDLAMRIMWIYNFAFKRPIPF (SEQ ID NO: 63);

SAFDLAMRIMWIYVFAWKRPIPF (SEQ ID NO: 64);

>HB36.4 (Asp47Ser, Ala60Val, Asn64Lys)

MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAEAVLQAVYETESAFDLA
MRIMWIYVFAFKRPIPFPHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 65);
>HB36.1 (Asp47Ser)
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAEAVLQAVYETESAFDLA
MRIMWIYAFAFNRPIPFSHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 66);
>HB36.2 (Ala60Val)
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAEAVLQAVYETEDAFDLA
MRIMWIYVFAFNRPIPFSHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 67);
>HB36.3 (Asp47Ser, Ala60Val)
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAEAVLQAVYETESAFDLA
MRIMWIYVFAFNRPIPFSHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 68);
>HB36.4_s4_E03
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDDEAAAVLQAVYETNHAFDLA
MRIHWIYVFAFKRKIPFLHAQKLARRLLELKQAASSPLP (SEQ ID NO: 69);
>HB36.4_s4_E05
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAAAVLKAVYATNSAFDLA
MRIIWIYVFAYKRKIPFAHAQKLARRLLELKQAASSPLP (SEQ ID NO: 70);
>HB36.4_s4_E06
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDFEADKVLQAVYETNSAFDLAM
RINWIYVFAFKRPIPFVHAQKLARRLLELKQAASSPLP (SEQ ID NO: 71);
>HB36.4_s4_E07
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAAAVLKAVYETNSAFDLA
MRINWIYVFAFKRKIPFAHAQKLARRLLELKQAASSPLP (SEQ ID NO: 72);
>HB36.4_s4_E08
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEADKVLQAVYDTNSAFDLA
MTIHWIYNFAFKRKIPFLHAPKLARRLLELKLAASSPLP (SEQ ID NO: 73);
>HB36.4_s4_E09
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDDEADRVLQAVYETNSAFDLAM
RINWIYVFAFKRTIPFAHAQKLARRLLELKQAASSPLP (SEQ ID NO: 74);
>HB36.4_s4_E10
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDYEADKVLQAVYETNSAFDLA

MRIHWIYIFAFKRPIPFVHAQKLARRLLELKQAASSPLP (SEQ ID NO: 75);

>HB36.4_s4_E11

MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEADAVLKAVYETNSAFDLA MRIHWIYNFAFKRKIPFVHAQKLARRLLELKQAASSPLP (SEQ ID NO: 76);

>HB36.4_s4_E12

MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDDEADKVLQAVYATNSAFDLA MRIHWIYNFAYKRTIPFVHAQKLARRLLELKQAASSPLP (SEQ ID NO: 77);

>HB36.4_s4_E13

MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDDEAARVLKAVYATDSAFDLA MRIHWIYNFAFKRKIPFLHAQKLARRLLELKQAASSPLP (SEQ ID NO: 78);

>HB36.4_s4_E14

MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEADKVLQAVYATNSAFDLA MRIHWIYIFAFKRTIPFIHAQKLARRLLELKQAASSPLP (SEQ ID NO: 79);

>HB36.4_s4_E17

MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDYEADEVLKAVYATNSAFDLA MRIHWIYNFAFKRKIPFTHAQKLARRLLELKQAASSPLP (SEQ ID NO: 80);

>HB36.4_s4_E18

MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAAKVLQAVYETNSAFDLA MKIHWIYNFAFKRTIPFVHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 81); and

>HB36.4_s4_E19

MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEADKVLQAVYATNSAFDLA MKIHWIYIFAFKRTIPFIHAQKLARRLLELKQAASSPLP (SEQ ID NO: 82).

In various preferred embodiments, HB36.4 (SAFDLAMRIMWIYVFAF (SEQ ID NO: 7)) is modified such that one or more of the following is true: R1 is His; R7 is Lys; R9 is Tyr, Asn, or His; R13 is Phe, Leu, Thr, or Asn; and R16 is Trp. In another embodiment, R10 is Trp. In a further embodiment, R2 and/or R5 is Ala. In a further embodiment, R17 is Phe.

As will be appreciated by those of skill in the art, these are just exemplary polypeptides falling under the scope of the claim. The table below provides per position allowable substitutions on an HB36.4 scaffold.

HB36.4:

(1) Central helix recognition motif from Serine 47- Phenylalanine 63 (SAFDLAMRIMWIYVFAF (SEQ ID NO: 7)); Also Phe 69 outside of that recognition motif (MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAEAVLQAVYETE**SAFDL AMRIMWIYVFAF**KRPIP**F**PHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 65))

(2) Allowable positions were determined from yeast display selections of HB36.4 variants to SC1918/H1 HA coupled to deep sequencing (see attached for further details). The threshold was no more than 80% depletion in the frequency of a given mutant in the selection library after two selection sorts by FACS. Positions listed in bold font indicate positions that make contact with the HA surface.

Table 1. Allowable substitutions on an HB36.4 scaffold

Position	**HB36.4 Residue**	**Allowable**
47 R1	Ser	ala,phe,his,lys,met,asn,gln,thr,val,tyr, asp
48 R2	Ala	All Amino Acids
49	**Phe**	**Phe**
50 R3	**Asp**	**Ala,Glu,Gly,Asn,Pro,Ser,Tyr**
51 R4	Leu	Phe
52 R5	Ala	All amino acids
53 R6	**Met**	**Phe,His,Ile,Leu,Gln,Thr**
54 R7	Arg	gly,lys,gln,thr
55 R8	Ile	asn,gln,val,trp
56 R9	**Met**	**Gly,Ile,Lys,Leu,Asn,Arg,Ser,Thr,Val,Tyr, His**
57 R10	**Trp**	**Phe**
58 R11	Ile	phe,ser,thr,val
59 R12	Tyr	cys,asp,phe,his,asn,ser
60 R13	**Val**	**Ala,Phe,Ile,Leu,Asn,Gln,Thr,Tyr**
61 R14	**Phe**	**Glu,Leu**
62 R15	Ala	gly,lys,arg,ser
63 R16	Phe	cys,his,lys,leu,met,asn,gln,arg,thr,val,trp,tyr
69 R17	**Phe**	**Tyr**

The table below shows where single point mutants from HB36.4 (SAFDLAMRIMWIYVFAF (SEQ ID NO: 7)) are shown to result in increased binding affinity. Thus, in other embodiments, the polypeptide comprises amino acid substitutions relative to HB36.4 as follows (singly or in combination):

Table 2. HB36.4 point mutations that show increased binding affinity

Position	HB36.4 Residue	Increased Affinity
47 R1	Ser	His
54 R7	Arg	Lys
56 R9	Met	His, Asn, Tyr
60 R13	Val	Phe, Leu, Thr, Asn
63 R16	Phe	Trp

All of these embodiments can be combined with any other embodiment, unless the context clearly dictates otherwise.

In a second aspect, the present invention provides polypeptides comprising an amino acid sequence according to general formula II

R1-R2-R3-R4-R5-R6-R7-R8-R9-Ala-R10-R11-Phe (SEQ ID NO: 83), wherein

R1 is selected from the group consisting of Phe and Val;

R2 is selected from the group consisting of Ser, Ala, Phe, Gly, Ile, Lys, Leu, Met, Asn, Pro, Gln, Arg, Thr, and Val;

R3 is selected from the group consisting of Glu, and Asp;

R4 is selected from the group consisting of Asn, His, Ile, Lys, Leu, Met, Arg, Ser, and Thr;

R5 is selected from the group consisting of Leu, Phe, Ile, Met, Asn, Gln, and Val;

R6 is selected from the group consisting of Ala, Asp, Lys, Met, Asn, Gln, Arg, Glu, and Val;

R7 is selected from the group consisting of Phe, Asp, Asn, and Tyr;

R8 is selected from the group consisting of Glu, Ala, Asp, Gly, His, Lys, Leu, Met, Asn, Gln, Arg, Ser, Thr, Val, and Trp;

R9 is selected from the group consisting of Leu, Phe, Ile, Met, and Val;

R10 is selected from the group consisting of Leu, Ile, Met, and Tyr; and

R11 is selected from the group consisting of Ser, Ala, Gly, and Tyr;

In one embodiment, general formula II is R1-R2-R3-R4-R5-R6-R7-R8-R9-Ala-R10-R11-Phe-**X1-R12-R13-X2-R14** (SEQ ID NO: 84), wherein R1 through R11 are as defined above, and wherein

X1 is 5-15 amino acids in length, wherein each position can be any amino acid;

R12 is selected from the group consisting of Gln, Tyr, Phe, Met, Arg, Lys, and Gly;

R13 is selected from the group consisting of Tyr, Asp, Met, Asn, and Ser;

X2 is any amino acid; and

R14 is selected from the group consisting of Ser, Arg, and Lys.

In various embodiments, X1 is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 amino acids in length. In another embodiment, X1 comprises the amino acid sequence TNKDTPDRW-**Z1**-KVA (SEQ ID NO: 85) where Z1 is Ala, Lys, Arg, Gly, or Thr.

In another embodiment, that can be combined with any other embodiments herein, general formula II is **A1**- R1-R2-R3-R4-R5-R6-R7-R8-R9-Ala-R10-R11-Phe-**X1-R12-R13-X2-R14-B1** (SEQ ID NO: 86), wherein R1 through R14 and X1 are as defined above, wherein A1 and/or B1 are optionally present, and wherein:

A1 comprises the amino acid sequence: Z1-ASTRGSGRPW-Z2 (SEQ ID NO: 87), wherein Z1 is absent or is Met, and Z2 is selected from group consisting of Gly, Arg, Lys, Asp and

B1 comprises the amino acid sequence G-Z1-TPEEVKKHYE (SEQ ID NO: 88), where Z1 is R or K

The inventors have discovered that polypeptides comprising the amino acid sequence of general formula II (derived from HB80.3, as described in more detail in the attached) form helices that recognize and are strong binders to Influenza A hemagglutinin. Thus, the polypeptides can be used, for example, to treat and/or limit development of an influenza infection.

In one embodiment, the polypeptide comprises the peptide FSENLAFELALSF (SEQ ID NO: 89), or a variant including 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more variant positions of FSENLAFELALSF (SEQ ID NO: 89) according to general formula II. In other embodiments, the polypeptide comprises amino acid substitutions relative to HB80.3 as follows (singly or in combination)

Position	HB80.3 Residue	Increased Affinity
12 R-1	Gly	Lys/Arg
14 R2	Ser	Lys/Arg
17 R5	Leu	Val/Ile
18 R6	Ala	Thr/Lys

21 R9	Leu	Ile
24 R12	Ser	Tyr
39	Gln	Arg/Tyr
42	Ser	Lys/Arg

In other exemplary embodiments, the polypeptide comprises or consists of a polypeptide selected from the group consisting of

FSENLAFELALA (SEQ ID NO: 90);

>HB80.3_s4_E81: FSENVAFEIALSF (SEQ ID NO: 91);

>HB80.3_s4_E82: FSENVAFEIALSF (SEQ ID NO: 92);

>HB80.3_s4_E83: FRENIAFEIALYF (SEQ ID NO: 93);

>HB80.3_s4_E84: FSENVAFEIALSF (SEQ ID NO: 94);

>HB80.3_s4_E85: FSENIAFELALYF (SEQ ID NO: 95);

>HB80.3_s4_E86: FSENVAFELALYF (SEQ ID NO: 96);

>HB80.3_s4_E87: FSENIAFELALYF (SEQ ID NO: 97);

>HB80.3_s4_E88: FKENLEFEIALSF (SEQ ID NO: 98);

>HB80.3_s4_E89: FSENVAFEIALSF (SEQ ID NO: 99);

>HB80.3_s4_E90: FSENVAFELALYF (SEQ ID NO: 100);

>HB80.3_s4_E91: FSENVAFELALYF (SEQ ID NO: 101);

>HB80.3_s4_E92: FSENVAFEIALSF (SEQ ID NO: 102);

>HB80.3_s4_E93: FSENVAFELALYF (SEQ ID NO: 103);

>HB80.3_s4_E94: FSENVAFELALYF (SEQ ID NO: 104);

>HB80.3_s4_E95: FSENVAFELALYF (SEQ ID NO: 105);

>HB80.3_s4_E96: FSENVAFEIALSF (SEQ ID NO: 106);

>HB80.3_s4_E97: FSENVAFEIALSF (SEQ ID NO: 107);

>HB80.3_s4_E98: FSENVAFEIALSF (SEQ ID NO: 108);

>HB80.3_s4_E99: FSENLAFELALYF (SEQ ID NO: 109);

>HB80.3_s4_E100: FSENVAFEIALSF (SEQ ID NO: 110);

>HB80.3_s5_E01: FSENVAFEIALSF (SEQ ID NO: 111);

>HB80.3_s5_E04: FSENVAFEIALSF (SEQ ID NO: 112);

>HB80.3_02: FSENIAFEIALSF (SEQ ID NO: 113);

>HB80.3_16: FSENIAFEIALSF (SEQ ID NO: 114);

>HB80.3 (Asp12Gly, Ala24Ser, Met26Thr, Asn36Lys, Delta54-95)
FSENLAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 115);
>HB80.3_s4_E81: FSENVAFEIALSFTNKDTPDRWKKVARYVR (SEQ ID NO: 116);
>HB80.3_s4_E82: FSENVAFEIALSFTNKDTPDRWAKVARYVR (SEQ ID NO: 117);
>HB80.3_s4_E83: FRENIAFEIALYFTNKDTPDRWRKVARYVK (SEQ ID NO: 118);
>HB80.3_s4_E84: FSENVAFEIALSFTNKDTPDRWRKVARYVR (SEQ ID NO: 119);
>HB80.3_s4_E85: FSENIAFELALYFTNKDTPDRWGKVARYVR (SEQ ID NO: 120);
>HB80.3_s4_E86: FSENVAFELALYFTNKDTPDRWKKVARYVK (SEQ ID NO: 121);
>HB80.3_s4_E87: FSENIAFELALYFTNKDTPDRWKKVARYVK (SEQ ID NO: 122);
>HB80.3_s4_E88: FKENLEFEIALSFTNKDTPDRWKKVAYYVR (SEQ ID NO: 123);
>HB80.3_s4_E89: FSENVAFEIALSFTNKDTPDRWRKVARYVR (SEQ ID NO: 124);
>HB80.3_s4_E90: FSENVAFELALYFTNKDTPDRWTKVARYVK (SEQ ID NO: 125);
>HB80.3_s4_E91: FSENVAFELALYFTNKDTPDRWTKVARYVK (SEQ ID NO: 126);
>HB80.3_s4_E92: FSENVAFEIALSFTNKDTPDRWRKVARYVR (SEQ ID NO: 127);
>HB80.3_s4_E93: FSENVAFELALYFTNKDTPDRWGKVAQYVR (SEQ ID NO: 128);
>HB80.3_s4_E94: FSENVAFELALYFTNKDTPDRWAKVARYVK (SEQ ID NO: 129);
>HB80.3_s4_E95: FSENVAFELALYFTNKDTPDRWTKVARYVK (SEQ ID NO: 130);
>HB80.3_s4_E96: FSENVAFEIALSFTNKDTPDRWRKVAYYVR (SEQ ID NO: 131);
>HB80.3_s4_E97: FSENVAFEIALSFTNKDTPDRWRKVARYVR (SEQ ID NO: 132);
>HB80.3_s4_E98: FSENVAFEIALSFTNKDTPDRWAKVARYVR (SEQ ID NO: 133);
>HB80.3_s4_E99: FSENLAFELALYFTNKDTPDRWAKVAYYVK (SEQ ID NO: 134);
>HB80.3_s4_E100: FSENVAFEIALSFTNKDTPDRWKKVARYVK (SEQ ID NO: 135);
>HB80.3_s5_E01: FSENVAFEIALSFTNKDTPDRWRKVARYVR (SEQ ID NO: 136);
>HB80.3_s5_E04: FSENVAFEIALSFTNKDTPDRWRKVARYVR (SEQ ID NO: 137);
>HB80.3_02: FSENIAFEIALSFTNKDTPDRWKKVAQYVK (SEQ ID NO: 138);
>HB80.3_16: FSENIAFEIALSFTNKDTPDRWKKVAQYVK (SEQ ID NO: 139);
FAENLAFELALSF (SEQ ID NO: 141);
FGENLAFELALSF (SEQ ID NO: 142);
FIENLAFELALSF (SEQ ID NO: 143);
FKENLAFELALSF (SEQ ID NO: 144);
FRENLAFELALSF (SEQ ID NO: 145);

FTENLAFELALSF (SEQ ID NO: 146);
FVENLAFELALSF (SEQ ID NO: 147);
FSENIAFELALSF (SEQ ID NO: 148);
FSENVAFELALSF (SEQ ID NO: 149);
FSENLKFELALSF (SEQ ID NO: 150);
FSENLRFELALSF (SEQ ID NO: 151);
FSENLTFELALSF (SEQ ID NO: 152);
FSENLAFSLALSF (SEQ ID NO: 153);
FSENLAFELALYF (SEQ ID NO: 154);
FSENLAFELALSFTNKDTPDRWAKVARYVS (SEQ ID NO: 156);
FSENLAFELALSFTNKDTPDRWAKVAYYVS (SEQ ID NO: 157);
FSENLAFELALSFTNKDTPDRWAKVAQYVK (SEQ ID NO: 158);
FSENLAFELALSFTNKDTPDRWAKVAQYVR (SEQ ID NO: 159);
FSENLAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 160);
FAENLAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 161);
FGENLAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 162);
FIENLAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 163);
FKENLAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 164);
FRENLAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 165);
FTENLAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 166);
FVENLAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 167);
FSENIAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 168);
FSENVAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 169);
FSENLKFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 170);
FSENLRFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 171);
FSENLTFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 172);
FSENLAFSLALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 173);
FSENLAFELALYFTNKDTPDRWAKVAQYVS (SEQ ID NO: 174);
FSENLAFELALSFTNKDTPDRWAKVAQYVS (SEQ ID NO: 175);
FSENLAFELALSFTNKDTPDRWAKVARYVS (SEQ ID NO: 176);
FSENLAFELALSFTNKDTPDRWAKVAYYVS (SEQ ID NO: 177);

FSENLAFELALSFTNKDTPDRWAKVAQYVK (SEQ ID NO: 178);

FSENLAFELALSFTNKDTPDRWAKVAQYVR (SEQ ID NO: 179);

>HB80 Met26Thr

MASTRGSGRPWDFSENLAFELALAFTNKDTPDRWANVAQYVSGRTPEEVKKHY EILVEDIKYIESGKVPFPNYRTTGGNMKTDEKRFRNLKIRLE (SEQ ID NO: 180);

>HB80 Asn36Lys

MASTRGSGRPWDFSENLAFELALAFMNKDTPDRWAKVAQYVSGRTPEEVKKH YEILVEDIKYIESGKVPFPNYRTTGGNMKTDEKRFRNLKIRLE (SEQ ID NO: 181);

>HB80.1 (Met26Thr, Asn36Lys)

MASTRGSGRPWDFSENLAFELALAFTNKDTPDRWAKVAQYVSGRTPEEVKKHY EILVEDIKYIESGKVPFPNYRTTGGNMKTDEKRFRNLKIRLE (SEQ ID NO: 182);

>HB80.2 (Met26Thr, Asn36Lys, Delta54-95)

MASTRGSGRPWDFSENLAFELALAFTNKDTPDRWAKVAQYVSGRTPEEVKKHYE (SEQ ID NO: 183);

>HB80.3 (Asp12Gly, Ala24Ser, Met26Thr, Asn36Lys, Delta54-95)

MASTRGSGRPWGFSENLAFELALSFTNKDTPDRWAKVAQYVSGRTPEEVKKHYE (SEQ ID NO: 184);

MASTRGSGRPWKFSENLAFELALSFTNKDTPDRWAKVAQYVSGRTPEEVKKHYE (SEQ ID NO: 185);

MASTRGSGRPWRFSENLAFELALSFTNKDTPDRWAKVAQYVSGRTPEEVKKHYE (SEQ ID NO: 186);

>HB80.3_s4_E81

MASTRGSGRPWRFSENVAFEIALSFTNKDTPDRWKKVARYVRGRTPEEVKKHYE (SEQ ID NO: 187);

>HB80.3_s4_E82

MASTRGSGRPWKFSENVAFEIALSFTNKDTPDRWAKVARYVRGRTPEEVKKHYE (SEQ ID NO: 188);

>HB80.3_s4_E83

MASTRGSGRPWGFRENIAFEIALYFTNKDTPDRWRKVARYVKGRTPEEVKKHYE (SEQ ID NO: 189);

>HB80.3_s4_E84

MASTRGSGRPWRFSENVAFEIALSFTNKDTPDRWRKVARYVRGRTPEEVKKHYE (SEQ ID NO: 190);

>HB80.3_s4_E85

MASTRGSGRPWGFSENIAFELALYFTNKDTPDRWGKVARYVRGRTPEEVKKHYE (SEQ ID NO: 191);

>HB80.3_s4_E86

MASTRGSGRPWKFSENVAFELALYFTNKDTPDRWKKVARYVKGRTPEEVKKHYE (SEQ ID NO: 192);

>HB80.3_s4_E87

MASTRGSGRPWKFSENIAFELALYFTNKDTPDRWKKVARYVKGRTPEEVKKHYE (SEQ ID NO: 193);

>HB80.3_s4_E88

MASTRGSGRPWKFKENLEFEIALSFTNKDTPDRWKKVAYYVRGRTPEEVKKHYE (SEQ ID NO: 194);

>HB80.3_s4_E90

MASTRGSGRPWKFSENVAFELALYFTNKDTPDRWTKVARYVKGRTPEEVKKHYE (SEQ ID NO: 196);

>HB80.3_s4_E92

MASTRGSGRPWKFSENVAFEIALSFTNKDTPDRWRKVARYVRGRTPEEVKKHYE (SEQ ID NO: 198);

>HB80.3_s4_E93

MASTRGSGRPWKFSENVAFELALYFTNKDTPDRWGKVAQYVRGRTPEEVKKHYE (SEQ ID NO: 199);

>HB80.3_s4_E94

ASTRGSGRPWKFSENVAFELALYFTNKDTPDRWAKVARYVKGRTPEEVKKHYE (SEQ ID NO: 200);

>HB80.3_s4_E96

MASTRGSGRPWKFSENVAFEIALSFTNKDTPDRWRKVAYYVRGRTPEEVKKHYE (SEQ ID NO: 202);

>HB80.3_s4_E98

MASTRGSGRPWRFSENVAFEIALSFTNKDTPDRWAKVARYVRGRTPEEVKKHYE (SEQ

ID NO: 204);

>HB80.3_s4_E99

MASTRGSGRPWKFSENLAFELALYFTNKDTPDRWAKVAYYVKGRTPEEVKKHYE (SEQ ID NO: 205);

>HB80.3_s4_E100

MASTRGSGRPWRFSENVAFEIALSFTNKDTPDRWKKVARYVKGRTPEEVKKHYE (SEQ ID NO: 206);

>HB80.3_s5_E01

MASTKGSGKPWKFSENVAFEIALSFTNKDTPDRWRKVARYVRGKTPEEVKKHYE (SEQ ID NO: 207); and

>HB80.3_02

MASTRGSGRPWKFSENIAFEIALSFTNKDTPDRWKKVAQYVKGRTPEEVKKHYE (SEQ ID NO: 209).

As will be appreciated by those of skill in the art, these are just exemplary polypeptides falling under the scope of the claim. The table below provides per position allowable substitutions on an HB80.3 scaffold.

(1) Central helix recognition motif from Phenylalanine 13- Phenylalanine 25; Also Tyrosine 40 that is outside of that recognition motif.

(MASTRGSGRPWG**FSENLAFELALSF**TNKDTPDRWAKVAQ**Y**VSGRTPEEVKKHYE (SEQ ID NO: 184))

Allowable positions were determined from yeast display selections of HB80.3 variants to SC1918/H1 HA coupled to deep sequencing (see attached for further details). The threshold was no more than 80% depletion in the frequency of a given mutant in the selection library after two selection sorts by FACS. Positions listed in bold font indicate positions that make contact with the HA surface.

Table 3. Allowable substitutions on an HB80.3 scaffold

Position	**HB80.3 Residue**	**Allowable**
13 R1	**Phe**	**Val**
14 R2	**Ser**	**Ala,Phe,Gly,Ile,Lys,Leu,Met,Asn,Pro,Gln,Arg,Thr,Val**
15 R3	Glu	Asp
16 R4	Asn	His,Ile,Lys,Leu,Met,Arg,Ser,Thr
17 R5	**Leu**	**Phe,Ile,Met,Asn,Gln,Val**
18 R6	**Ala**	**Asp,Lys,Met,Asn,Gln,Arg,Val**

19 R7	Phe	Asp,Asn,Tyr
20 R8	Glu	Ala,Asp,Gly,His,Lys,Leu,Met,Asn,Gln,Arg,Ser,Thr,Val,Trp
21 R9	**Leu**	**Phe,Ile,Met,Val**
22	Ala	Ala
23 R10	Leu	Ile,Met,Tyr
24 R11	Ser	Ala,Gly,Tyr
25	**Phe**	**Phe**
39 **R12**	Gln	Tyr,Phe,Met,Arg,Lys,Gly
40 **R13**	Tyr	Asp,Met,Asn,Ser
42 **R14**	Ser	Arg,Lys

The table below shows where single point mutants from HB80.3 are shown to result in increased binding affinity. Thus, in other embodiments, the polypeptide comprises amino acid substitutions relative to HB80.3 as follows (singly or in combination).

Table 4. HB80.3 point mutations that show increased binding affinity

Position	**HB80.3 Residue**	**Increased Affinity**
14 R2	Ser	Ala, Gly,Ile,Lys,Arg,Thr,Val
17 R5	Leu	Ile,Val
18 R6	Ala	Lys,Arg
20 R8	Glu	Ser
21 R9	Leu	Ile
24 R11	Ser	Tyr

In various preferred embodiments, HB80.3 (FSENLAFELALSF (SEQ ID NO: 89)) is modified such that one or more of the following is true: R2 is Ala, Gly, Ile, Lys, Arg, Thr, or Val; R5 is Ile or Val; R6 is Lys or Arg; R8 is Ser; R9 is Ile; and/or R11 is Tyr.

All of these embodiments can be combined with any other embodiment, unless the context clearly dictates otherwise.

In a third aspect, the invention provides polypeptides comprising or consisting of a polypeptide selected from the group consisting of

>HB3

MADTLLILGDSLSAGYQMLAEFAWPFLLNKKWSKTSVVNASISGDTSQQGLARLPALL KQHQPRWVLVELGGNDGLEGFQPQQTEQTLRQILQDVKAANAEPLLMQIRPPANYGRR

YNEAFSAIYPKLAKEFDVPLLPFFMEEVYLKPQWMQDDGIHPNYEAQPFIADWMAKQL QPLVNH (SEQ ID NO: 155);

>HB54

MAETKNFTDLVEATKWGNSLIKSAKYSSKDKMAIYNYTKNSSPINTPLRSANGDVNKLS ENIQEQVRQLDSTISKSVTPDSVYVYRLLNLDYLSSITGFTREDLHMLQQTNEGQYNSKL VLWLDFLMSNRIYRENGYSSTQLVSGAALAGRPIELKLELPKGTKAAYIDSKELTAYPG QQEVLLPRGTEYAVGTVELSKSSQKIIITAVVFKK (SEQ ID NO: 140); and

>HB78

MFTGVIIKQGCLLKQGHTRKNWSVRKFILREDPAYLHYYYPLGYFSPLGAIHLRGCVVT SVESEENLFEIITADEVHYFLQAATPKERTEWIKAIQMASR (SEQ ID NO: 211).

Each of these polypeptides form helices that recognize and are strong binders to Influenza A hemagglutinin. Thus, the polypeptides can be used, for example, to treat and/or limit development of an influenza infection

In a fourth aspect, the present invention provides a polypeptide comprising or consisting of any helix coming from a peptide or a protein that docks and binds against the HA epitope recognized by the polypeptides of the invention. In one embodiment, the helix is 15-17 residues in length, similar to the HB36.4 and HB80.3 helices disclosed above

As used throughout the present application, the term "polypeptide" is used in its broadest sense to refer to a sequence of subunit amino acids. The polypeptides of the invention may comprise L-amino acids, D-amino acids (which are resistant to L-amino acid-specific proteases in vivo), or a combination of D- and L-amino acids. The polypeptides described herein may be chemically synthesized or recombinantly expressed. The polypeptides may be linked to other compounds to promote an increased half-life in vivo, such as by PEGylation, HESylation, PASylation, glycosylation, or may be produced as an Fc-fusion or in deimmunized variants. Such linkage can be covalent or non-covalent as is understood by those of skill in the art.

In a further embodiment, the polypeptides of any embodiment of any aspect of the invention may further comprise a tag, such as a detectable moiety or therapeutic agent. The tag(s) can be linked to the polypeptide through covalent bonding, including, but not limited to, disulfide bonding, hydrogen bonding, electrostatic bonding, recombinant fusion and conformational bonding. Alternatively, the tag(s) can be linked to the polypeptide by means of

one or more linking compounds. Techniques for conjugating tags to polypeptides are well known to the skilled artisan. Polypeptides comprising a detectable tag can be used diagnostically to, for example, assess if a subject has been infected with influenza virus or monitor the development or progression of an influenza virus infection as part of a clinical testing procedure to, e.g., determine the efficacy of a given treatment regimen. However, they may also be used for other detection and/or analytical and/or diagnostic purposes. Any suitable detection tag can be used, including but not limited to enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, radioactive materials, positron emitting metals, and nonradioactive paramagnetic metal ions. The tag used will depend on the specific detection/analysis/diagnosis techniques and/or methods used such as immunohistochemical staining of (tissue) samples, flow cytometric detection, scanning laser cytometric detection, fluorescent immunoassays, enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), bioassays (e.g., neutralization assays), Western blotting applications, etc. For immunohistochemical staining of tissue samples preferred tags are enzymes that catalyze production and local deposition of a detectable product. Enzymes typically conjugated to polypeptides to permit their immunohistochemical visualization are well known and include, but are not limited to, acetylcholinesterase, alkaline phosphatase, beta-galactosidase, glucose oxidase, horseradish peroxidase, and urease. Typical substrates for production and deposition of visually detectable products are also well known to the skilled person in the art. The polypeptides can be labeled using colloidal gold or they can be labeled with radioisotopes, such as ^{33}P, ^{32}P, ^{35}S, ^{3}H, and ^{125}I. Polypeptides of the invention can be attached to radionuclides directly or indirectly via a chelating agent by methods well known in the art.

When the polypeptides of the invention are used for flow cytometric detections, scanning laser cytometric detections, or fluorescent immunoassays, the tag may comprise, for example, a fluorophore. A wide variety of fluorophores useful for fluorescently labeling the polypeptides of the invention are known to the skilled artisan. When the polypeptides are used for in vivo diagnostic use, the tag can comprise, for example, magnetic resonance imaging (MRI) contrast agents, such as gadolinium diethylenetriaminepentaacetic acid, to ultrasound contrast agents or to X-ray contrast agents, or by radioisotopic labeling.

The polypeptides of the invention can also be attached to solid supports, which are particularly useful for in vitro assays or purification of influenza virus or HA protein. Such solid

supports might be porous or nonporous, planar or nonplanar and include, but are not limited to, glass, cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene supports. The polypeptides can also, for example, usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of affinity chromatography. They can also usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction. The microspheres can be used for isolation of influenza virus or HA protein from a sample containing influenza virus or HA protein. As another example, the polypeptides of the invention can usefully be attached to the surface of a microtiter plate for ELISA.

The polypeptides of the invention can be fused to marker sequences to facilitate purification. Examples include, but are not limited to, the hexa-histidine tag, the myc tag or the flag tag.

The polypeptides of the invention can be conjugated to an antigen recognized by the immune system of a subject to which the polypeptide is administered. Conjugation methods for attaching the antigens and polypeptide are well known in the art and include, but are not limited to, the use of cross-linking agents. The polypeptide will bind to the influenza virus HA protein and the antigen will initiate a T-cell attack on the conjugate that will facilitate destruction of the influenza virus.

In another embodiment of any aspect herein, the present invention provides retro-inverso polypeptides corresponding to the polypeptides of the invention. Retro-inverso polypeptides of the invention comprise or consist of D-amino acids assembled in a reverse order from that of L-sequence polypeptide versions of the polypeptides disclosed above, thus maintaining the overall topology of the polypeptide, and maintaining HA binding.

In a fifth aspect, the present invention provides isolated nucleic acids encoding a polypeptide of the present invention. The isolated nucleic acid sequence may comprise RNA or DNA. As used herein, "isolated nucleic acids" are those that have been removed from their normal surrounding nucleic acid sequences in the genome or in cDNA sequences. Such isolated nucleic acid sequences may comprise additional sequences useful for promoting expression and/or purification of the encoded protein, including but not limited to polyA sequences, modified Kozak sequences, and sequences encoding epitope tags, export signals, and secretory signals, nuclear localization signals, and plasma membrane localization signals. It will be

apparent to those of skill in the art, based on the teachings herein, what nucleic acid sequences will encode the polypeptides of the invention.

In a sixth aspect, the present invention provides recombinant expression vectors comprising the isolated nucleic acid of any aspect of the invention operatively linked to a suitable control sequence. "Recombinant expression vector" includes vectors that operatively link a nucleic acid coding region or gene to any control sequences capable of effecting expression of the gene product. "Control sequences" operably linked to the nucleic acid sequences of the invention are nucleic acid sequences capable of effecting the expression of the nucleic acid molecules. The control sequences need not be contiguous with the nucleic acid sequences, so long as they function to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the nucleic acid sequences and the promoter sequence can still be considered "operably linked" to the coding sequence. Other such control sequences include, but are not limited to, polyadenylation signals, termination signals, and ribosome binding sites. Such expression vectors can be of any type known in the art, including but not limited plasmid and viral-based expression vectors. The control sequence used to drive expression of the disclosed nucleic acid sequences in a mammalian system may be constitutive (driven by any of a variety of promoters, including but not limited to, CMV, SV40, RSV, actin, EF) or inducible (driven by any of a number of inducible promoters including, but not limited to, tetracycline, ecdysone, steroid-responsive). The construction of expression vectors for use in transfecting prokaryotic cells is also well known in the art, and thus can be accomplished via standard techniques. (See, for example, Sambrook, Fritsch, and Maniatis, in: Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989; *Gene Transfer and Expression Protocols*, pp. 109-128, ed. E.J. Murray, The Humana Press Inc., Clifton, N.J.), and the Ambion 1998 Catalog (Ambion, Austin, TX). The expression vector must be replicable in the host organisms either as an episome or by integration into host chromosomal DNA. In a preferred embodiment, the expression vector comprises a plasmid. However, the invention is intended to include other expression vectors that serve equivalent functions, such as viral vectors.

In a seventh aspect, the present invention provides host cells that have been transfected with the recombinant expression vectors disclosed herein, wherein the host cells can be either prokaryotic or eukaryotic. The cells can be transiently or stably transfected. Such transfection

of expression vectors into prokaryotic and eukaryotic cells can be accomplished via any technique known in the art, including but not limited to standard bacterial transformations, calcium phosphate co-precipitation, electroporation, or liposome mediated-, DEAE dextran mediated-, polycationic mediated-, or viral mediated transfection. (See, for example, *Molecular Cloning: A Laboratory Manual* (Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press; *Culture of Animal Cells: A Manual of Basic Technique, 2nd Ed.* (R.I. Freshney. 1987. Liss, Inc. New York, NY). A method of producing a polypeptide according to the invention is an additional part of the invention. The method comprises the steps of (a) culturing a host according to this aspect of the invention under conditions conducive to the expression of the polypeptide, and (b) optionally, recovering the expressed polypeptide. The expressed polypeptide can be recovered from the cell free extract, but preferably they are recovered from the culture medium. Methods to recover polypeptide from cell free extracts or culture medium are well known to the man skilled in the art.

In an eighth aspect, the present invention provides antibodies that selectively bind to the polypeptides of the invention. The antibodies can be polyclonal, monoclonal antibodies,human ized antibodies, and fragments thereof, and can be made using techniques known to those of skill in the art . As used herein, "selectively bind" means preferential binding of the antibody to the polypeptide of the invention, as opposed to one or more other biological molecules, structures, cells, tissues, etc., as is well understood by those of skill in the art.

In a ninth aspect, the present invention provides pharmaceutical compositions, comprising one or more polypeptides of the invention and a pharmaceutically acceptable carrier. The pharmaceutical compositions of the invention can be used, for example, in the methods of the invention described below. The pharmaceutical composition may comprise in addition to the polypeptide of the invention (a) a lyoprotectant; (b) a surfactant; (c) a bulking agent; (d) a tonicity adjusting agent; (e) a stabilizer; (f) a preservative and/or (g) a buffer.
In some embodiments, the buffer in the pharmaceutical composition is a Tris buffer, a histidine buffer, a phosphate buffer, a citrate buffer or an acetate buffer. The pharmaceutical composition may also include a lyoprotectant, e.g. sucrose, sorbitol or trehalose. In certain embodiments, the pharmaceutical composition includes a preservative e.g. benzalkonium chloride, benzethonium, chlorohexidine, phenol, m-cresol, benzyl alcohol, methylparaben, propylparaben, chlorobutanol, o-cresol, p-cresol, chlorocresol, phenylmercuric nitrate, thimerosal, benzoic acid, and various

mixtures thereof. In other embodiments, the pharmaceutical composition includes a bulking agent, like glycine. In yet other embodiments, the pharmaceutical composition includes a surfactant e.g., polysorbate-20, polysorbate-40, polysorbate- 60, polysorbate-65, polysorbate-80 polysorbate-85, poloxamer-188, sorbitan monolaurate, sorbitan monopalmitate, sorbitan monostearate, sorbitan monooleate, sorbitan trilaurate, sorbitan tristearate, sorbitan trioleaste, or a combination thereof. The pharmaceutical composition may also include a tonicity adjusting agent, e.g., a compound that renders the formulation substantially isotonic or isoosmotic with human blood. Exemplary tonicity adjusting agents include sucrose, sorbitol, glycine, methionine, mannitol, dextrose, inositol, sodium chloride, arginine and arginine hydrochloride. In other embodiments, the pharmaceutical composition additionally includes a stabilizer, e.g., a molecule which, when combined with a protein of interest substantially prevents or reduces chemical and/or physical instability of the protein of interest in lyophilized or liquid form. Exemplary stabilizers include sucrose, sorbitol, glycine, inositol, sodium chloride, methionine, arginine, and arginine hydrochloride.

The polypeptides may be the sole active agent in the pharmaceutical composition, or the composition may further comprise one or more other active agents suitable for an intended use, including but not limited to anti-HA and anti-NA antibodies.

In a tenth aspect, the present invention provides methods for treating and/or limiting an influenza infection, comprising administering to a subject in need thereof a therapeutically effective amount of one or more polypeptides of the invention, salts thereof, conjugates thereof, or pharmaceutical compositions thereof, to treat and/or limit the influenza infection. When the method comprises treating an influenza infection, the one or more polypeptides are administered to a subject that has already been infected with the influenza virus, and/or who is suffering from symptoms (including but not limited to chills, fever, sore throat, muscle pains, coughing, weakness, fatigue, and general discomfort) indicating that the subject is likely to have been infected with the influenza virus. As used herein, "treat" or "treating" means accomplishing one or more of the following: (a) reducing influenza viral titer in the subject; (b) limiting any increase of influenza viral titer in the subject; (c) reducing the severity of flu symptoms; (d) limiting or preventing development of flu symptoms after infection; (e) inhibiting worsening of flu symptoms; (f) limiting or preventing recurrence of flu symptoms in subjects that were previously symptomatic for influenza infection.

When the method comprises limiting an influenza infection, the one or more polypeptides are administered prophylactically to a subject that is not known to have been infected, but may be at risk of exposure to the influenza virus. As used herein, "limiting" means to limit influenza infection in subjects at risk of influenza infection. Given the nature of seasonal influenza outbreaks, virtually all subjects are at risk of exposure, at least at certain times of the year. Groups at particularly high risk include children under age 18, adults over the age of 65, and individuals suffering from one or more of asthma, diabetes, heart disease, or any type of immunodeficiency.

The methods of the invention can be used to treat any individual infected with influenza virus, including but not limited to influenza virus A, influenza virus B, and influenza virus C. The methods are preferably used to treat influenza A virus infections caused by influenza A viruses of phylogenetic group I, in particular comprising HA of the H1 or H5 subtype.

As used herein, a "therapeutically effective amount" refers to an amount of the polypeptide that is effective for treating and/or limiting influenza infection. The polypeptides are typically formulated as a pharmaceutical composition, such as those disclosed above, and can be administered via any suitable route, including orally, parentally, by inhalation spray, rectally, or topically in dosage unit formulations containing conventional pharmaceutically acceptable carriers, adjuvants, and vehicles. The term parenteral as used herein includes, subcutaneous, intravenous, intra-arterial, intramuscular, intrasternal, intratendinous, intraspinal, intracranial, intrathoracic, infusion techniques or intraperitoneally. Dosage regimens can be adjusted to provide the optimum desired response (e.g., a therapeutic or prophylactic response). A suitable dosage range may, for instance, be 0.1 ug/kg-100 mg/kg body weight; alternatively, it may be 0.5 ug/kg to 50 mg/kg; 1 ug/kg to 25 mg/kg, or 5 ug/kg to 10 mg/kg body weight. The polypeptides can be delivered in a single bolus, or may be administered more than once (e.g., 2, 3, 4, 5, or more times) as determined by an attending physician.

In certain embodiments, the polypeptides of the invention neutralize influenza virus infectivity. While not being limited by any mechanism of action, neutralizing activity may be achieved by inhibiting fusion of the influenza virus and the membrane of the targeted cell, including a membrane of an intracellular compartment, such as an endosome. The polypeptides of the invention were designed to target an HA epitope that is absent in HA post-conformational change. Since the HA protein conformational change leads to fusion of the viral and cell

membrane, polypeptide binding to the HA protein in its pre-fusion form may prevent fusion. In various embodiments, the polypeptides of the invention prevent influenza virus from infecting host cells by at least 99%, at least 95%, at least 90%, at least 85%, at least 80%, at least 75%, at least 70%, at least 60%, at least 50%, at least 45%, at least 40%, at least 45%, at least 35%, at least 30%, at least 25%, at least 20%, or at least 10% relative to infection of host cells by influenza virus in the absence of the polypeptides. Neutralization can, for instance, be measured as described in "Laboratory techniques in influenza," edited by F. -X. Meslin, M. M. Kaplan and H. Koprowski (1996), 4th edition, Chapters 15-17, World Health Organization, Geneva.

The polypeptides according to the invention can bind to the HA protein with any suitable affinity constant (K_d value) that provides therapeutic or prophylactic benefit. In various embodiments, the K_d value is lower than $0.2*10^{-4}$ M, $1.0*10^{-5}$M, $1.0*10^{-6}$M, $1.0*10^{-7}$M, $1.0*10^{-8}$M, $1.0*10^{-9}$M, $1.0*10^{-10}$M, $1.0*10^{-11}$M, or $1.0*10^{-12}$M. Affinity constants can for instance be measured using surface plasmon resonance, i.e., an optical phenomenon that allows for the analysis of real-time biospecific interactions by detection of alterations in protein concentrations within a biosensor matrix, for example, using the BIACORE system (Pharmacia Biosensor AB, Uppsala, Sweden).

In a eleventh aspect, the present invention provides methods for diagnosing an influenza infection, or monitoring progression of an influenza infection, comprising

(a) contacting a biological sample from a subject suspected of having an influenza infection with a diagnostically effective amount of one or more polypeptides of the invention under conditions suitable for binding of the polypeptide to a viral HA protein present in the sample;

(b) removing unbound polypeptide and/or sample; and

(c) detecting polypeptide-viral HA binding complexes,

where the presence of such binding complexes indicates that the subject has an influenza infection, or provides a measure progression of an influenza infection.

The methods of this aspect of the invention can be used to more accurately identify patients that may be suffering from an influenza infection and to thus provide more informed determination of treatment options by an attending caregiver. Individuals at risk of an influenza infection are as described above. The methods can also be used to monitor progression of an influenza infection; in this embodiment, the subject is known to be infected, and the methods can

be used, for example, as a data point for an attending caregiver to determine whether to initiate, modify, or continue a particular course of therapy, such as treatment with neuraminidase or M2 protein inhibitors.

The biological sample may be any suitable biological sample including, but not limited to blood, serum, nasal secretions, tissue or other biological material from a subject at risk of infection.

The sample may first be manipulated to make it more suitable for the method of detection. "Manipulation" includes, but is not limited to treating the the sample in such a way that any influenza virus in the sample will disintegrate into antigenic components such as proteins, polypeptides or other antigenic fragments. The polypeptides of the invention are contacted with the sample under conditions which allow the formation of an complex between the human polypeptides and influenza virus or antigenic components thereof that may be present in the sample. The formation of such complexes, if any, indicating the presence of influenza virus in the sample, is then detected and measured by suitable means. Such methods include, but are not limited to homogeneous and heterogeneous binding immunoassays, such as radioimmunoassays (RIA), ELISA, immunofluorescence, immunohistochemistry, FACS, BIACORE and Western blot analyses. Suitable conditions to promote binding of the test compounds to one or more polypeptide of the invention can be determined by those of skill in the art, based on the teachings herein.

The polypeptides of the invention for use in this aspect may comprise a conjugate as disclosed above, to provide a tag useful for any detection technique suitable for a given assay. The tag used will depend on the specific detection/analysis/diagnosis techniques and/or methods used. The methods may be carried in solution, or the polypeptide(s) of the invention may be bound or attached to a carrier or substrate, e.g., microtiter plates (ex: for ELISA), membranes and beads, etc. Carriers or substrates may be made of glass, plastic (e.g., polystyrene), polysaccharides, nylon, nitrocellulose, or teflon, etc. The surface of such supports may be solid or porous and of any convenient shape. In one embodiment, conditions are selected to identify test compounds that bind to the polypeptide of the invention with a K_d value lower than $0.2*10^{-4}$ M, $1.0*10^{-5}$M, $1.0*10^{-6}$M, $1.0*10^{-7}$M, $1.0*10^{-8}$M, $1.0*10^{-9}$M, $1.0*10^{-10}$M, $1.0*10^{-11}$M, or $1.0*10^{-12}$M.

In a twelfth aspect, the present invention provides methods for identifying candidate

influenza vaccines, comprising

(a) contacting test compounds with a polypeptide of the present invention under conditions suitable for polypeptide binding; and

(b) identifying those test compounds that bind to the polypeptide of the invention, wherein such test compounds are candidate influenza vaccines.

As discussed above, the polypeptides of the present invention were designed to target an HA epitope that is absent in HA post-conformational change. Thus, the polypeptides of the invention can be viewed as specific binders to an HA epitope, similar to antibody binding to a specific epitope. Vaccines can be produced, for example, by selecting small molecules (ie: mimotopes) that bind to an antibody specific to a viral epitope. Thus, the present methods involve substituting one or more polypeptides of the present invention for the antibody in such assay to identify candidate influenza vaccines.

Suitable conditions to promote binding of the test compounds to one or more polypeptide of the invention can be determined by those of skill in the art, based on the teachings herein. The polypeptides of the invention for use in this aspect may comprise a conjugate as disclosed above, to provide a tag useful for any detection technique suitable for a given assay. The tag used will depend on the specific detection/analysis/diagnosis techniques and/or methods used, as discussed above. The methods may be carried in solution, or the polypeptide(s) of the invention may be bound or attached to a carrier or substrate, as discussed above. Based on the teachings herein, it is within the level of skill in the art to determine specific conditions for the various types of diagnostic assays disclosed in this aspect of the invention. In one embodiment, conditions are selected to identify test compounds that bind to the polypeptide of the invention with a K_d value lower than $0.2*10^{-4}$ M, $1.0*10^{-5}$M, $1.0*10^{-6}$M, $1.0*10^{-7}$M, $1.0*10^{-8}$M, $1.0*10^{-9}$M, $1.0*10^{-10}$M, $1.0*10^{-11}$M, or $1.0*10^{-12}$M.

When the test compounds comprise polypeptide sequences, such polypeptides may be chemically synthesized or recombinantly expressed. Recombinant expression can be accomplished using standard methods in the art, as disclosed above. Such expression vectors can comprise bacterial or viral expression vectors, and such host cells can be prokaryotic or eukaryotic. Synthetic polypeptides, prepared using the well-known techniques of solid phase, liquid phase, or peptide condensation techniques, or any combination thereof, can include natural and unnatural amino acids. Amino acids used for peptide synthesis may be standard Boc (Nα-

amino protected Nα-t-butyloxycarbonyl) amino acid resin with standard deprotecting, neutralization, coupling and wash protocols, or standard base-labile Nα-amino protected 9-fluorenylmethoxycarbonyl (Fmoc) amino acids. Both Fmoc and Boc Nα-amino protected amino acids can be obtained from Sigma, Cambridge Research Biochemical, or other chemical companies familiar to those skilled in the art. In addition, the polypeptides can be synthesized with other Nα-protecting groups that are familiar to those skilled in this art. Solid phase peptide synthesis may be accomplished by techniques familiar to those in the art and provided, such as by using automated synthesizers.

When the test compounds comprise antibodies, such antibodies can be polyclonal or monoclonal. The antibodies can be humanized, fully human, or murine forms of the antibodies. Such antibodies can be made by well-known methods, such as described in Harlow and Lane, Antibodies; A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., (1988).

When the test compounds comprise nucleic acid sequences, such nucleic acids may be produced by any suitable means, such as chemical synthesis. The nucleic acids may be DNA or RNA, and may be single stranded or double. Similarly, such nucleic acids can be chemically or enzymatically synthesized by manual or automated reactions, using standard techniques in the art. If synthesized chemically or by in vitro enzymatic synthesis, the nucleic acid may be purified prior to introduction into the cell. For example, the nucleic acids can be purified from a mixture by extraction with a solvent or resin, precipitation, electrophoresis, chromatography, or a combination thereof. Alternatively, the nucleic acids may be used with no or a minimum of purification to avoid losses due to sample processing.

When the test compounds comprise compounds other than polypeptides, antibodies, or nucleic acids, such compounds can be made by any of the variety of methods in the art for conducting organic chemical synthesis.

In a thirteenth aspect, the present invention provides methods for identifying candidate compounds for treating, limiting, and/or diagnosing influenza infection, comprising

(a) contacting an influenza HA protein with (i) test compounds and (ii) a polypeptide of the present invention, under conditions suitable for binding of the HA protein to the polypeptide of the present invention; and

(b) identifying those test compounds that outcompete the polypeptide for binding to the HA protein, wherein such test compounds are candidate compounds for treating, limiting, and/or diagnosing influenza infection.

In this aspect, the methods identify test compounds that compete with the polypeptides of the invention for binding to HA, and thus such candidate compounds may be useful in any of the other methods of the invention disclosed herein. Any suitable test compound can be used, as disclosed above in the eleventh aspect of the invention.

In general, competitive inhibition is measured by means of an assay, wherein an HA composition is admixed with the polypeptide(s) of the invention and the test compounds to be screened. In one embodiment, the test compounds to be screened are present in excess. Protocols based upon ELISAs are suitable for use in such competition studies. In certain embodiments, one may pre-mix the polypeptide(s) of the invention with varying amounts of test compounds to be screened (e.g., 1:10, 1:20, 1:30, 1:40, 1:50, 1:60, 1:70, 1:80, 1:90 or 1:100) for a period of time prior to applying to the HA composition. In other embodiments, the polypeptide(s) of the invention and varying amounts of test compounds to be screened are admixed during exposure to the HA composition. Any suitable detection means can be used binding. In one embodiment, the polypeptide(s) of the invention are tagged for detection, as discussed above. In this embodiment, the detectable label will decrease in the presence of competitive test compounds. The reactivity of the (labeled) polypeptide of the invention in the absence of test compound could serve as one suitable control. Preferably, competitive test compounds will, when present in excess, inhibit specific binding of the polypeptide(s) of the invention to HA by at least 10%, preferably by at least 25%, more preferably by at least 50%, and most preferably by at least 75% to 90% or even greater.

Exemplary conditions for HA binding studies can be carried out as disclosed in the examples that follow.

All of these aspects/embodiments disclosed herein can be combined with any other aspect/embodiment, unless the context clearly dictates otherwise.

Example 1. Design of proteins for binding to influenza hemagglutinin

Abstract

We describe a general computational method for designing proteins that bind a surface patch of interest on a target macromolecule. Favorable interactions between disembodied amino-acid residues and the target surface are identified and used to anchor *de novo* designed interfaces. The method was used to design proteins that bind a conserved surface patch on the stem of the influenza hemagglutinin (HA) from the 1918 H1N1 pandemic virus. After affinity maturation, two of the designed proteins, HB36 and HB80, bind H1 and H5 HAs with low-nanomolar affinity. Further, HB80 inhibits the HA fusogenic conformational changes induced at low pH. The crystal structure of HB36 in complex with 1918/H1 HA revealed that the actual binding interface is nearly identical to that in the computational design model. Such designed proteins may be useful for both diagnostics and therapeutics.

Introduction

Molecular recognition is central to biology, and high-affinity binding proteins, such as antibodies, are invaluable for both diagnostics and therapeutics (*1*). Current methods for producing antibodies and other proteins that bind a protein of interest involve screening of large numbers of variants generated by the immune system or by library construction (*2*). The computer-based design of high-affinity binding proteins is a fundamental test of the current understanding of the physical-chemical basis of molecular recognition and, if successful, would be a powerful complement to current library-based screening methods since it would allow targeting of specific patches on a protein surface. Recent advances in computational design of protein interactions have yielded switches in interaction specificity (*3*), methods to generate modest-affinity complexes (*4, 5*), two-sided design of a novel protein interface (*6*), and design of a high-affinity interaction by grafting known key residues onto an unrelated protein scaffold (*7*). However, the capability to target an arbitrarily selected protein surface has remained elusive.

Influenza presents a serious public-health challenge and new therapies are needed to combat viruses that are resistant to existing antivirals (*8*) or escape neutralization by the immune system. Hemagglutinin (HA) is a prime candidate for drug development as it is the major player in viral invasion of cells lining the respiratory tract. While most antibodies bind to the rapidly

varying head region of HA, recently two antibodies, CR6261 and F10, were structurally characterized (*9, 10*) that bind to a region on the HA stem, which is conserved among all group 1 influenza strains (*11*). Here, we describe a computational method for designing protein-protein interactions *de novo*, and use the method to design high-affinity binders to the conserved stem region on influenza HA.

Computational design method

In devising the computational design strategy, we considered features common to dissociable protein complexes. During protein complex formation, proteins bury on average ~1,600Å^2 of solvent-exposed surface area (*12*). Interfaces typically contain several residues that make highly optimized van der Waals, hydrogen bonding, and electrostatic interactions with the partner protein; these interaction hotspots contribute a large fraction of the binding energy (*13*).

Our strategy thus centers on the design of interfaces that have both high shape complementarity and a core region of highly optimized, hotspot-like residue interactions. We engineer high-affinity interactions and high shape complementarity into scaffold proteins in two steps (see Fig. 1): (i) disembodied amino-acid residues are computationally docked or positioned against the target surface to identify energetically favorable configurations with the target surface; and (ii) shape-complementary configurations of scaffold proteins are computed that incorporate the key residues.

Design of HA-binding proteins

The surface on the stem of HA recognized by neutralizing antibodies consists of a hydrophobic groove that is flanked by two loops that place severe steric constraints on binding to the epitope (Fig. 2A-B) (*14*). In the first step of our design protocol (Fig. 1), the disembodied residues found through computational docking cluster into three regions (HS1, HS2, and HS3; Fig. 1). In HS1, a Phe side chain forms an energetically favorable aromatic-stacking interaction with Trp21 on chain 2 of the HA (HA2) (HA residue numbering corresponds to the H3 subtype sequence-numbering convention). In HS2, the nonpolar residues Ile, Leu, Met, Phe, and Val, make favorable van der Waals interactions with both the hydrophobic groove and HS1 (Fig. 1). In HS3, a Tyr side chain forms a hydrogen bond to Asp18 on HA2 and van der Waals interactions with the A-helix on HA2. The Tyr in HS3 resembles the conformation of a Tyr

residue observed on the antibody in the structure of the HA and CR6261 Fab complex ; the HS1 and HS2 interactions are not found in the antibody structures (*9, 10, 15*) .

In the second step, we searched a set of 865 protein structures selected for ease of experimental manipulation (*16*) for scaffolds capable of supporting the disembodied hotspot residues and shape complementary to the stem region. Each scaffold protein was docked against the stem region using the feature-matching algorithm PatchDock™ (*17*), identifying hundreds of compatible binding modes for each scaffold (260,000 in total). These coarse-grained binding modes were then refined using RosettaDock™ (*18*) with a potential function that favored configurations that maximized the compatibility of the scaffold protein backbone with as many hotspot residues as possible. Next, residues from the hotspot-residue libraries were incorporated on the scaffold. First, for each Phe conformation in HS1, scaffold residues with backbone atoms within 4Å of the hotspot residue were identified. For each of these candidate positions, the scaffold protein was placed to coincide with the backbone of the hotspot, the residue was modeled explicitly, and the rigid-body orientation was minimized. If no steric clashes were observed and the Phe was in contact with Trp21 and Thr41 of HA2 (Fig. 2B), the placement of the first hotspot was deemed successful; otherwise, another HS1 Phe conformation was selected and the process was repeated. For each success with HS1, nonpolar residues were incorporated at positions in the scaffold protein, from which the HS2 interactions could be realized, and the remainder of the scaffold protein surface was then redesigned using RosettaDesign™ (*19*).

Designing proteins also containing HS3 interactions was more challenging due to the large number of combinations of residue placements to be considered. To generate designs containing all three hotspot regions, we started by superimposing the scaffold protein on the backbone of the Tyr residue in HS3 (as for the Phe HS1 residue above). We then searched for two positions on the scaffold protein that were nearest to residues in HS1 and HS2 and were best aligned to them. These positions were then simultaneously designed to Phe in the case of HS1 and to nonpolar residues in the case of HS2. RosettaDesign™ (*19*) was then used to redesign the remainder of the interface on the scaffold protein, allowing sequence changes within a distance of 10 Å of the HA.

Experimental and structural characterization

A total 51 designs using the two hotspot-residue concept and 37 using the three-residue concept were selected for testing. The designs are derived from 79 different protein scaffolds and differ from the scaffold by on average 11 mutations. Genes encoding the designs were synthesized, cloned into a yeast-display vector, and transformed into yeast strain EBY100 (*20, 21*). Upon induction, the designed protein is displayed on the cell surface as a fusion between an adhesion subunit of the Aga2p yeast protein and a C-terminal c-myc tag. Cells expressing designs were incubated with 1 uM of biotinylated SC1918/H1 (A/South Carolina/1/1918 (H1N1)) HA ectodomain, washed, and dual-labeled with phycoerythrin-conjugated streptavidin and fluorescein-conjugated anti-c-myc antibody. Binding was measured by flow cytometry with the two fluorescent tags allowing simultaneous interrogation of binding to HA and surface display of the design.

73 designs were surface-displayed, and 2 showed reproducible binding activity towards the HA stem region (*22*)(for models, see Fig. 2C-F). One design, HA Binder 36 (HB36) used the two-residue hotspot, and bound to the HA with an apparent dissociation constant (K_d) of 200 nM(*23*)(Fig. 2G, **Fig. 6** . The starting scaffold, Structural Genomics target APC36109, a protein of unknown function from *B. stearothermophilus* (PDB entry 1U84), did not bind HA (Fig. 6), indicating that binding is mediated by the designed surface on HB36. A second design, HB80, used the three-residue hotspot and bound HA only weakly (Fig. 2H). The scaffold from which this design was derived, the MYB domain of the RAD transcription factor from *A. Majus* (PDB code: 2CJJ) (*24*), again did not bind the HA (Fig. 7).

In the computational models of the two designs (Fig. 2C-F), the hotspot residues are buttressed by a concentric arrangement of hydrophobic residues with an outer ring of polar and charged residues as often observed in native protein-protein interfaces. Both designs present a row of hydrophobic residues on a helix that fits into the HA hydrophobic groove. The complexes each bury approximately 1,550Å^2 surface area (total), close to the mean value for dissociable protein interactions (*12*) and slightly larger than the total surface area buried by each of the two neutralizing antibodies (*9, 10*). The helical binding modes in these designs are very different from the loop-based binding observed in the antibody-bound structures.

Affinity maturation

The computational design protocol is far from perfect; the energy function that guides design contains numerous approximations (*25*) and conformational sampling is incomplete. We used affinity maturation to identify shortcomings in the design protocol. Libraries of HB36 and HB80 variants were generated by single site-saturation mutagenesis at the interface, or by error-prone PCR (epPCR), and subjected to two rounds of selection for binding to HA using yeast-surface display (*21, 24*).

For both designed binders, the selections converged on a small number of substitutions that increase affinity and provide insight into how to improve the underlying energy function. Among the key contributions to the energetics of macromolecular interactions are short-range repulsive interactions due to atomic overlaps, electrostatic interactions between charged and polar atoms, and the elimination of favorable interactions with solvent (desolvation). The affinity-increasing substitutions point to how each of these contributions can be better modeled in the initial design calculations.

Repulsive interactions: For HB36, substitution of Ala60 with the isosteres Thr/Val increased the apparent binding affinity 25-fold (apparent K_d's for all design variants are listed in Table 5).

Table 5. Dissociation constants (K_d) for binding of design variants to SC1918 HA

Design	K_d [nM]*
1U84 (HB36 Scaffold)	NB (NB)
HB36	200 (>2000)
HB36 Asp47Ser	5
HA36 Ala60Val	8
HB36.3 (HB36 Asp47Ser, Ala60Val)	4 (29)
HB36.4 (HB36 Asp47Ser, Ala60Val, Asn64Lys)	4 (22)
2CJJ (HB80 Scaffold)	NB
HB80	>5000
HB80 Met26Thr	100
HB80 Asn36Lys	300
HB80 Met26Thr Asn36Lys	7.5
HB80 Δ54-95, Met26Thr, Asn36Lys	5
HB80.3 (HB80 Δ54-95, Asp12Gly, Ala24Ser, Met26Thr, Asn36Lys)	3 (38)
* K_d was determined using yeast surface display titrations.	
Number in parentheses indicates K_d determined by SPR.	
NB, no binding.	

These substitutions fill a void between the designed protein and the HA surface, but were not included in the original design because they were disfavored by steric clashes within HB36 (Fig. 3A). Backbone minimization, however, readily relieved these clashes resulting in higher predicted affinity for the substitutions. For HB80, a Met26Thr mutation significantly increased binding compared to the starting design. Modeling showed that Met26 disfavored the conformation of the Tyr hotspot residue, rationalizing the substitution to a smaller residue (Fig. 3B). More direct incorporation of backbone minimization in the design algorithm should allow identification of such favorable interactions from the start, whereas insuring that hotspot residues are fully relaxed in the design would eliminate unfavorable interactions.

Electrostatics: In HB36, the substitution to Lys at position 64 places a complementary charge adjacent to an acidic pocket on HA near the conserved stem region (Fig. 3C); in HB80, an Asn36Lys substitution positions a positive charge 6.5Å from the negative Asp18 on HA2 (Fig. 3D). These substitutions all enhance electrostatic complementarity in the complex. The lysines

were not selected in the design calculations because the magnitudes of surface-electrostatic interactions between atoms outside of hydrogen-bonding range are largely reduced; improvement of the electrostatic model would evidently allow design of higher-affinity binders from the start.

Desolvation: In HB36, 8 different substitutions at Asp47 increased apparent affinity by over an order of magnitude compared to the original design (Table 6); the highest-affinity substitution was Asp47Ser that increased binding affinity *circa* 40-fold. The design of an unfavorable charged group in this position likely stems from underestimation of the energetic cost of desolvating Asp47 by the aliphatic Ile18 on HA2 (Fig. 3E); the substitutions remedy this error by replacing the Asp with residues that are less costly to desolvate upon binding. In HB80, an Asp12Gly substitution relieves the desolvation by the neighboring Ile56 on HA2 (Fig. 3F). With improvements in the solvation model, the deleterious Asp residues would not be present in starting designs.

Table 6 Selected mutations at Asp47 of HB36 design that increased binding affinity >10-fold relative to original design.

Clone	Mutation(s)	Approx Binding Affinity*
C1	D47S	+++
C3	D47H	+++
C4	D47H,P70S	+++
D3	D47N,G7S	++
E1	D47Y,G19C	++
A2	D47L,P68L,P70L	++
A4	D47R,P70L	++
B6	D47W	++
B3	D47R	+
B2	D47E	+

*Approximate binding affinity by 5-pt yeast titration.
+++, K_d~2-5 nM ++ 5-15nM, + 15-40 nM

The favorable substitutions were combined and the proteins were expressed with a His-tag in *E. coli* and purified by nickel affinity and size-exclusion chromatography. The variant HB36.3, incorporating the Asp47Ser and Ala60Val substitutions, bound to SC1918/H1 HA as confirmed by surface plasmon resonance (SPR;), ELISA, and co-elution on a size-exclusion column (data not shown). The HB36.4 variant, which incorporates Asp47Ser, Ala60Val, and Asn64Lys, bound to SC1918/H1 HA with a dissociation constant measured by SPR of 22nM and

an off-rate of $7 \cdot 10^{-3}$ s^{-1} (Table 7). Co-incubation with an excess of CR6261 Fab abolished binding to the HA (Fig. 3G), consistent with HB36.4 binding in close proximity to the same stem epitope on the HA. For the HB80 design, the combination of the affinity-increasing mutations reduced surface expression on yeast, indicative of poor stability. Therefore, we excised a C-terminal stretch (Δ54-95) greatly boosting surface expression of the design with no significant loss of binding affinity (Fig. 8). HB80.3, which incorporates the truncation as well as the Asp12Gly, Ala24Ser, Met26Thr, and Asn36Lys substitutions, has a K_d=38nM with off-rate of $4 \cdot 10^{-2}$ s^{-1} by SPR. As with HB36.4, co-incubating HA with the CR6261 Fab completely abolished binding to HB80.3 (Fig. 3H), consistent with the designed binding mode.

Table 7 Affinity and kinetic binding constants for specified design variants. All measurements were recorded using surface plasmon resonance. Numbers in parentheses indicate error associated with the measurement.

Design Variant	K_d [nM]	k_{on} [$M^{-1}s^{-1}$]	k_{off} [s^{-1}]
HB36.3 (D47S, A60V)	29.0±0.6	1.2±0.1 e6	3.5±0.3 e2
HB36.4 (D47S, A60V,N64K)	22.3±0.9	3.2±0.2 e5	7±1 e3
HB80.3 (D12G,A24S,M26T,N36K)	38±2	1.0±0.2 e6	3.9±0.8 e2

Site-directed alanine mutagenesis of several core positions on each affinity-matured design partially or completely knocked out HA binding (Table 8 , Fig. 9) supporting the computational model of the designed interfaces (*26*). Furthermore, no mutations were uncovered during selection for higher affinity that were inconsistent with the designed binding modes.

Table 8 Summary of alanine scanning mutagenesis of key residues at the interface of HB36 and HA80. Binding was measured by yeast surface display titrations on two separate days. NB marks no binding at 1μM HA. ΔΔG was computed from the change in K_d relative to HB36.3 at the assay temperature of 294 K.

Construct	K_d [nM]	ΔΔG [kcal/mol]
HB36.3 (D47S,A60V)	5.0±0.5	--
HB36.3 F49A	NB	>3.4
HB36.3 M53A	115±35	1.8±0.2
HB36.3 W57A	NB	>3.4
H80.1 (M26T, N36K)	7.5±1.0	--
HB80.1 F13A	NB	>2.9
HB80.1 F25A	NB	>2.9
HB80.1 Y40A	140±20	1.7±0.2

Crystal structure of the HB36.3-SC1918 HA complex

The crystal structure of HB36.3 in complex with the SC1918 HA ectodomain was determined to 3.1Å resolution. After molecular replacement using only the 1918/H1 HA structure as the search model (approximately 86% of the protein mass in the crystal asymmetric unit), clear electron density was observed for HB36.3 near the target surface in the HA stem region into which HB36.3 could be unambiguously placed. The orientation was essentially identical to the designed binding mode, with the modified surface of the main recognition helix packed in the hydrophobic groove on HA (Fig. 4A). To obtain unbiased density for the designed side chains, the native structure from which HB36.3 was derived (PDB entry: 1U86) was manually fit into the electron-density maps and contact side chains were pruned back to their β-carbon. After crystallographic refinement, electron density became apparent for the side chains of most of the contact residues on HB36.3, allowing the predominant rotamers to be assigned for Phe49, Trp57, Phe61, and Phe69. This unbiased density clearly shows that these four hydrophobic side chains are all positioned as in the designed model (Fig. 4B). The Met53 side chain is consistent with the design model (Fig. 4C), although other rotamers could also be fit to the map. For Met56, only very weak side-chain density was observed. Overall, the crystal structure is in excellent agreement with the designed interface, with no significant deviations at any of the contact positions.

Given the quite low (2 out of the 73 surface displayed proteins) design success rate and starting affinities, the atomic-level agreement between the designed and experimentally determined HB36.3-SC1918 HA complex is very encouraging and suggests that, despite their shortcomings, the current energy function and design methodology capture essential features of protein-protein interactions.

Cross-reactivity and inhibitory activity

The surface contacted by HB36.3 is accessible and highly conserved in the HAs of most group 1 influenza viruses, suggesting that it may be capable of binding not only other H1 HAs, but also other HA subtypes. Indeed, binding of HB36.3 to A/South Carolina/1/1918(H1N1) and A/WSN/1933(H1N1) is readily detectable in solution by gel filtration (data not shown), as well as high-affinity binding of HB36.4 to A/Vietnam/1203/2004 H5 subtype by yeast display (Fig. 10).

While a crystal structure of HB80 in complex with HA has not been obtained, the mutational data and the antibody-competition results suggest that HB80 also binds to the designed target surface, overlapping with HB36 and CR6261. Consequently, HB80.3 is also expected to be highly cross-reactive and binds with high affinity to A/Vietnam/1203/2004 H5 HA (Fig. 10), and to H1, H2, H5, and H6 subtypes by biolayer interferometry (Fig. 5 A,B). Overall, the pattern of HB80 binding mirrors that of CR6261 and binds most of the group 1 HAs tested, with no detectable binding to group 2 HAs.

Antibody CR6261 inhibits influenza virus replication by blocking the pH-induced refolding of HA, which drives fusion of the viral envelope with the endosomal membrane of the host cell. Given extensive overlap between the HB80.3 and CR6261 binding sites and its high affinity for SC1918 HA, it seemed plausible that HB80.3 would also block this conformational change. Indeed, HB80.3 inhibits the pH-induced conformational changes in both H1 and H5 HAs (Fig. 5C, Fig. 11)(*10*), suggesting that this design may possess virus-neutralizing activity against multiple influenza subtypes (*27*).

References and Notes for Example 1

1. H. Ledford, *Nature* **455**, 437 (2008).

2. R. A. Lerner, *Angew Chem Int Ed Engl* **45**, 8106 (2006).
3. T. Kortemme *et al.*, *Nat. Struct. Mol. Biol.* **11**, 371 (2004).
4. R. K. Jha *et al.*, *J Mol Biol* **400**, 257 (2010).
5. P. S. Huang, J. J. Love, S. L. Mayo, *Protein Sci* **16**, 2770 (2007).
6. J. Karanicolas *et al.*, *Mol. Cell* **in press**, (2011).
7. S. Liu *et al.*, *Proc Natl Acad Sci U S A* **104**, 5330 (2007).
8. E. Bautista *et al.*, *N Engl J Med* **362**, 1708 (2010).
9. J. Sui *et al.*, *Nat Struct Mol Biol* **16**, 265 (2009).
10. D. C. Ekiert *et al.*, *Science* **324**, 246 (2009).
11. Group 1 includes 10 of the 16 HA subtypes: H1, H2, H5, H6, H8, H9, H11, H12, H13, and H16. Group 2 includes the remaining 6 subtypes: H3, H4, H7, H10, H14, and H15.
12. L. Lo Conte, C. Chothia, J. Janin, *J Mol Biol* **285**, 2177 (1999).
13. T. Clackson, J. A. Wells, *Science* **267**, 383 (1995).
14. M. G. Rossmann, *J Biol Chem* **264**, 14587 (1989).
15. The other hotspot residues (HS1 and HS2) differed from the sidechains observed in the crystal structures in their conformation or identity. Each hotspot residue was further diversified by constructing all conformations, the terminal atoms of which coincided with those modeled above. For instance, for HS3, these consisted of all Tyr conformations that matched the position of the aromatic ring and hydrogen bond. This diversification step produced a 'fan' of backbone positions for each residue in the hotspot libraries.
16. Proteins in the scaffold set contained no disulfides, were expressed in E. coli, and were predicted to form monomers (see Supplemental Information).
17. D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H. J. Wolfson, *Nucleic Acids Res* **33**, W363 (2005).
18. J. J. Gray *et al.*, *J Mol Biol* **331**, 281 (2003).
19. B. Kuhlman *et al.*, *Science* **302**, 1364 (2003).
20. J. Chen, J. J. Skehel, D. C. Wiley, *Proc Natl Acad Sci U S A* **96**, 8967 (1999).
21. G. Chao *et al.*, *Nat Protoc* **1**, 755 (2006).
22. A third design HB35 bound HA at apparent low μM affinity; however, binding was only partially abolished upon co-incubation of HA with the CR6261 Fab, indicating of at most partial contact with the target surface on the stem region of HA, and so this design was

eliminated from further consideration. A handful of other designs bound HA albeit weakly and with incomplete reproducibility.

23. We recorded dissociation constants using two main methods: by titration of HA against yeast surface-displayed designs, and by fitting both kinetic and equilibrium measurements using surface plasmon resonance. As there is a discrepancy in determining Kd's between the methods, measurements derived from yeast surface-display titrations are listed as apparent Kd and should be viewed qualitatively. .

24. C. E. Stevenson *et al.*, *Proteins* **65**, 1041 (2006).

25. R. Das, D. Baker, *Annu Rev Biochem* **77**, 363 (2008).

26. The alanine-scan mutations were as follows: for HB36.3, Phe49, Met53, and Trp57; for HB80.1 Phe13, Phe25, and Tyr40 (**Table S4** and supplemental resutls).

27. HB36.4 was not able to block the pH-induced conformational changes in the H1 HA under identical assay conditions, even though HB36.4 and HB80.3 have very similar dissociation constants and kinetic off-rates at pH 7.5 (**Fig. 11**).

28. Computational designs were generated on resources generously provided by participants of Rosetta @ Home and the Argonne National Leadership Computing Facility. X-ray diffraction datasets were collected at the Stanford Synchrotron Radiation Lightsource beamline 9-2 and at the Advanced Photon Source beamline 23ID-B (GM/CA-CAT). Coordinates and structure factors were deposited in the Protein Data Bank (PDB) as entry 3R2X.

Supporting Material

Computational Design Methodology

Figure 1 provides a flowchart overview of the approach. This method is a generalization of a recently described approach for two-sided design of pairs of interacting proteins (S*1*). In that method surfaces of an ankyrin-repeat protein and a target protein were simultaneously mutated to introduce a hotspot region buttressed by a periphery of compatible interactions. The hotspot region in that method comprised aromatic residues that formed intermolecular hydrogen bonds. Our approach does not make any assumptions about the nature of the hotspot or the scaffold protein. We generate a hotspot region consisting of high-affinity interacting residues of all types and incorporate them into a variety of scaffold proteins. These generalizations allow us to design binders of potentially any protein surface.

Generating hotspot residues

Individual residues were docked against the target surface on influenza A/SC/1918/H1 hemagglutinin (hereafter referred to as HA) using RosettaDock™ (S*2*) starting from the structure of HA bound to the antibody fragment (Fab) CR6261(S*3*). We positioned the hydrophobic residues Leu, Val, Ile, Phe, Trp, Met, and Tyr against the surface of HA near Trp21 on HA2 (H3 HA sequencing numbering as in Protein Data Bank (PDB) entry 3GBN). Only conformations of the Phe were able to form satisfactory contacts with the surface, whereas the other residues either left small voids or buried polar atoms. Two dominant conformations of Phe were selected that were roughly 60° rotated relative to one another with respect to the center of the aromatic ring as hotspot residue 1 (HS1) (Fig. 1).

To compute the position of the second hotspot residue (HS2), we docked the same set of hydrophobic residues against the HA surface with the two major Phe conformations from HS1 placed to ensure that the residues that are selected form energetically favorable interactions with HA, as well as with HS1. This search yielded low-energy placements of Leu, Val, Ile, Phe, and Met for HS2.

Third, the Tyr, Asn, and Gln residues were docked against the HA2 A-helix region spanning Thr41 (Fig. 1) again including the Phe HS1 residues. We required each docked residue to form a hydrogen bond to the backbone carbonyl of Asp19 on HA2. Only a single dominant

orientation for a Tyr was identified that formed the requisite hydrogen bond, did not bury polar groups at the interface, and formed favorable van der Waals contacts with the A helix (Fig. 1).

All of the conformations identified by RosettaDock™ were diversified by generating inverse rotamers starting from their side-chain atoms nearest to the HA surface. These inverse rotamers were expanded to include rotamers one standard deviation away from the base rotamers in the Dunbrack library (S*4*) with the Rosetta™ commandline flags –ex1 –ex2.

A set of scaffold proteins

We selected a set of 865 proteins from the PDB in March 2009 according to the following criteria: they contained no disulfides, RNA, or DNA molecules, were solved by X-ray crystallography at a resolution better than 2.5Å, are reported to have been expressed in *E. coli*, are predicted to be monomeric by the Protein Quaternary Structure server(S*5*), and contain a single polypeptide chain of between 80 and 250 amino acids. The list was pruned at 70% sequence identity. Each structure was refined in the Rosetta™ forcefield by full side-chain repacking and minimization. .

Low-resolution docking of scaffold proteins against the target epitope

To obtain high shape-complementary configurations of the scaffold protein with respect to HA we used the PatchDock™ feature-matching algorithm (S*6*). Constraints were used to prune conformations of each scaffold protein that do not interact with Trp21 and Thr41 on HA2. The surviving conformations were clustered at 4Å root-mean-square deviation (RMSD). PatchDock™ was run with default parameters.

Backbone restraints

The hotspot-residue libraries are used to identify configurations of the scaffold protein with respect to HA that may accommodate the placement of these hotspot residues. Each hotspot residue computed in the library implies an approximate location for a position on the scaffold protein and an orientation for the $C\alpha$-$C\beta$ and the *C-N* vectors. For each hotspot residue h and each scaffold position i, we formulate scoring restraints R_i^h to bias conformational sampling to configurations that would favor the placement of the hotspot residues:

$$R_i^h = \min[0,(\Delta G_h + {}^{k}\!/_{n}(\vec{\beta}_i - \vec{\beta}_h)^2)[(\vec{\beta}_h - \vec{\alpha}_h)\bullet(\vec{\beta}_i - \vec{\alpha}_i)][(\vec{C}_h - \vec{N}_h)\bullet(\vec{C}_i - \vec{N}_i)]] \quad \text{(Eq. 1)}$$

where ΔG_h is the computed binding energy for hotspot residue h, is always negative and was chosen to be -3 in all design trajectories; β, α, C, and N, are the coordinates of the $C\beta$, $C\alpha$, C, and N atoms; k (the spring constant) is arbitrarily set to 0.5; min is the minimum function ensuring that the restraint is negative or zero; the quantities within the square brackets are the dot products of the relevant vectors; and $n = |\vec{\beta}_h - \vec{\alpha}_h||\vec{\beta}_i - \vec{\alpha}_i||\vec{C}_h - \vec{N}_h||\vec{C}_i - \vec{N}_i|$ is a normalization constant.

This form of the restraint function reaches a minimum when the distance between the $C\beta$ of the hotspot residue and a position on the scaffold is 0 and the $C\alpha$-$C\beta$ and C-N vectors are matched. Thus, a given restraint is best satisfied when a potential grafting position on the scaffold is perfectly aligned with a pre-computed hotspot residue. If the orientation of either of the two vectors of position i with respect to hotspot h is more than 90°, then R_i^h is set to 0. A library of n hotspot residues thus implies n restraints. Each residue i is then assigned the smallest of these n restraints:

(Eq. 2) $$R_i^{'} = \min_h(R_i^h)$$

Equation 2 then assigns the minimal restraint to each amino-acid position i on the scaffold, so that each scaffold position is affected only by the most appropriate hotspot restraint at any given time during conformational search.

Since only the locations of the $C\beta$ and the backbone atoms are required in evaluating Equation 2, the restraints can be computed efficiently during low-resolution Monte-Carlo based docking of the scaffold protein with respect to the HA surface. Importantly, the restraints can be used during minimization as Equation 1 is readily differentiable.

Hotspot-residue placement

We used two different protocols to design scaffolds that incorporate the computed hotspots. The more restrictive design strategy incorporated three hotspot residues (Tyr for HS3,

Phe for HS1, and a nonpolar residue for HS2); the less restrictive one incorporated two (Phe for HS1 and a nonpolar residue for HS2). We developed three methods for hotspot-residue placement for use in the different stages of design. Each starts with the configuration of the scaffold protein obtained from hotspot-residue guided docking and minimization with one of the hotspot-residue libraries. Except for Gly, Pro and disulfide-linked cysteines, interfacial residues on the scaffold protein within 10Å from the target protein were reduced to alanine to increase the chances of accommodating the hotspot residues.

Method 1: placement of the scaffold onto an idealized hotspot residue

The residues within the hotspot-residue libraries define configurations that are optimal for realizing the hotspot interaction. For a given interfacial scaffold position, we iterate over each of the nearby hotspot residues in the library and rotate and translate the scaffold protein so as to align it perfectly with the rotamer of the hotspot residue. Scaffold positions, for which the $C\beta$ atoms are farther than 4.0Å from the relevant hotspot residue or whose C-N or $C\alpha$-$C\beta$ vectors are misaligned with the hotspot residues by more than 60^{o}, are triaged to avoid compromising the initial, high shape complementary configuration of the two partners. We then minimize the rigid-body orientation and the side-chain degrees of freedom of the placed hotspot residue in a reduced forcefield that only considers the punitive energy terms for van der Waals clashes and rotameric energies. If the energy of the placed hotspot residue is higher than 1.0 Rosetta energy unit (R.e.u.), we discard this placement.

In the context of the two-residue hotspot designs, we used this strategy to place the hotspot residue Phe (HS1) on the scaffold proteins. In the case of the three-residue designs, we used this strategy to place the Tyr (HS3).

Method 2: placement of a hotspot residue onto a scaffold position

For each interfacial scaffold position, we minimize the configuration of the scaffold protein with respect to the target in the context of a single restraint (Eq. 1) derived from the hotspot residue. All other parameters and cutoffs are as in the previous section. We used this strategy to place HS2 in the two-residue hotspot designs.

Method 3: simultaneous placement of multiple hotspot residues

For each hotspot-residue library, we identify a position on the scaffold protein that produces the most favorable restraint score as defined by Equation 1 compared to the remainder of the hotspot-residue libraries. Each such scaffold position is then coupled to the appropriate hotspot-residue library. If not all hotspot-residue libraries are matched to different scaffold positions, the configuration of the scaffold with respect to the target is discarded. Upon success, we simultaneously redesign the identities of the relevant scaffold positions to those amino-acid identities contained in their matched hotspot-residue libraries. Since only a handful of positions are designed in this scheme and the identities of the designed residues are limited based on the relevant hotspot-residue library, the addition of off-rotameric conformations into the design step is computationally affordable. We used this scheme to place HS1 and HS2 in the three-residue hotspot designs.

Intensified conformational search in the design of scaffolds incorporating the three-residue hotspot

Preliminary trials using the three-residue placement approach (incorporating HS1-3) revealed that this combination of residues implies constraints on scaffold proteins that are very rarely met by proteins in the scaffold set. To increase the chances of identifying scaffolds that may incorporate the three-residue hotspot, we used a protocol that intensified the search in terms of both the backbone conformation of the scaffold proteins and their rigid-body orientations. This intensification was made possible by the computational-efficiency gains provided by the simultaneous-placement method.

For each scaffold, placement of the scaffold on the Tyr HS3 residue was attempted and was deemed successful if the Tyr hotspot residue's energy did not surpass 1R.e.u. and the Tyr formed a hydrogen bond with the Asp19 backbone carbonyl. We next conducted 4 trials of rigid-body docking followed by simultaneous placement (of HS1-2). During simultaneous hotspot-residue placement, backbone minimization and backrub (S7) were conducted to increase the chances of successful placement. In retrospect, backbone remodeling is likely to have contributed little to the success of the placement of the hotspot residues on HB80 as the backbone of this redesigned protein does not show significant differences from the starting wildtype structure.

Redesign of residues outside of the hotspot

Following the successful placement of residues from all hotspot-residue libraries, scaffold positions that are at most 10Å from the target protein are redesigned using RosettaDesign (S*8*), while the target protein side chains are allowed to repack. Gly, Pro and disulfide-linked cysteines are left as in the wildtype sequence. Three iterations of redesign and minimization were used to increase the likelihood that higher-affinity interactions are found, starting with a soft-repulsive potential, and gradually increasing the repulsive terms. The last design step uses the default all-atom forcefield with high weights on the steric clashes and rotameric strain to ensure that the designed residues do not assume high-energy conformations.

During these design simulations, the side chains of the placed hotspot residues are biased towards the coordinates of the idealized hotspot residues as present in the hotspot-residue library (similar to the implementation in ref. (S*9*)). This bias is implemented as harmonic coordinate restraints, typically on three atoms that define the functional group of the side chain, in effect pulling the placed hotspot residue's functional group towards its idealized position with respect to the target protein. For example, these atoms would be the three carbon atoms at the root of Tyr and Phe aromatic rings. To ensure that the placed residues are stable in their position on the scaffold, all restraints are gradually removed during the simulation and the last packing and minimization step is carried out in the absence of restraints.

Each resulting model is automatically filtered according to computed binding energy(S*10*), buried surface area, and shape complementarity (S*11*). Complexes that were predicted to have binding energies of more than -15R.e.u., surface areas of less than 1000Å^2, or shape-complementarity scores less than 0.65, were eliminated. At this stage, designs were reviewed manually, and a subset was selected for more rigorous evaluation. After the subsequently described modifications in the designs, some of the designs had statistics that failed these filters. While both HB36 (binding energy=-24, Sc=0.66, buried surface area=1620Å^2) and HB80 (binding energy=-19, Sc=0.72, buried surface area=1580Å^2) passed these filters, other designs with comparable statistics did not.

Minimizing the number of residue changes at the interface

For each design that passed the abovementioned filters, the contribution of each amino-acid substitution at the interface is assessed by singly reverting residues to their wild-type

identities and testing the effects of the reversion on the computed binding energy. If the difference in binding energy between the designed residue and the reverted one is less than 0.5R.e.u. in favor of the design, then the position is reverted to its wild-type identity.

A report of all residue changes was produced and each suggestion was reviewed manually. At this stage of manual review, additional mutations were introduced. These typically involve the introduction or removal of peripheral charges to better complement the charged surface of HA and did not routinely involve more than 5 substitutions per design.

An additional means of minimizing changes to the sequence of the original scaffold consisted of introducing sequence restraints during all stages of design. Briefly, mutations from the wildtype sequence were penalized according to their distance in the BLOSUM62 matrix (S*12*). The weight on these sequence restraints was set to 0.2.

Binding-energy calculations

In keeping with ref. (S*10*), the binding energy was defined as the difference between the total system energy in the bound and unbound states. In each state, interface residues were allowed to repack. For numerical stability, binding-energy calculations were repeated three times and the average taken.

Shape complementarity

Shape complementarity was computed using the CCP4 package v.6.0.2 (S*13*) using the *sc* program.

Experimental characterization

Expression and Purification of BirA

E. coli biotin ligase (BirA enzyme) was expressed and purified in a manner similar to previous reports(S*14*), but with an N-terminal His tag. The *birA* gene was amplified from an *E. coli* colony (wild-type strain MG1655) using primers DE389 (5'-agtcactaggtcatatgcatcaccatcaccatcacaaggataacaccgtgccactg-3' (SEQ ID NO: 195)) and DE390 (5'-agtcactaggtaagcttttatttttctgcactacgcagggatatttc-3'(SEQ I D NO: 197)). The PCR product was digested with *NdeI* and *HindIII* and ligated into similarly digested pET21a, yielding pDCE095. This vector was transformed into BL21(DE3) cells for protein expression.

BL21(DE3)/pDCE095 cells were grown in shake flasks in low salt LB medium at 37°C to an OD (600nm) of ~0.7, then shifted to 23°C and induced with the addition of IPTG (isopropyl-beta-D-thiogalactopyranoside) to a final concentration of 1mM. The culture was incubated at 23°C for ~16 hours after induction, then harvested by centrifugation (3000g, 10 minutes). The pellet from a 1L culture was resuspended in 50-100mL of lysis buffer (50mM Tris pH 8.0, 300mM potassium chloride, 10mM imidazole pH 8.0, with Roche EDTA-free protease inhibitor cocktail tablet) and the cells were lysed and homogenized by two passes through an EmulsiFlex™ C-3 cell disruptor (15kPSI). After clearing the lysates by centrifugation (25,000g, ~1 hour), the supernatant was incubated with NiNTA resin (Qiagen), washed with excess lysis buffer, and bound proteins were eluted (with 50mM Tris pH 8.0, 300mM potassium chloride, 250mM imidazole pH 8.0). After concentrating and buffer exchanging into 50mM potassium phosphate, pH6.5, 5% glycerol, 0.1mM dithiothreitol (DTT), the BirA was loaded onto a MonoQ column (GE Healthcare) and eluted with a linear gradient of 0-1M potassium chloride. BirA containing fractions were pooled, concentrated, and subjected to gel filtration. The final yield of BirA protein was approximately 10mg/L and >95% pure as assessed by SDS-PAGE. Purified BirA protein was concentrated to 5mg/mL in 50mM Tris, pH 7.5, 200mM potassium chloride, 5% glycerol, aliquoted, snap frozen in liquid nitrogen, and stored at -80°C.

Cloning, Expression and Purification of Hemagglutinins

Based on H3 numbering, cDNAs corresponding to residues 11-329 (HA1) and 1-176 (HA2) of the influenza A hemagglutinin (HA) were fused to an N-terminal gp67 signal peptide (amino acid sequence: MVLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFA (SEQ ID NO: 212)) and to a C-terminal trimerization domain and His-tag by overlap PCR, essentially as previously described(S*3*). The trimerization domain and His-tag were separated from the HA ectodomain by a thrombin cleavage site. For biotinylated HAs, a BirA target biotinylation site (amino-acid sequence: GGGLNDIFEAQKIEWHE (SEQ ID NO: 213)) was inserted between the HA and the thrombin site. The resulting PCR products were digested with SfiI, and inserted into a custom baculovirus transfer vector, pDCE198. Recombinant bacmids were generated using the Bac-to-Bac system (Invitrogen) and viruses were rescued by transfecting purified bacmid DNA into Sf9 cells using Cellfectin II (Invitrogen). HA proteins were produced by infecting

suspension cultures of Hi5 cells (Invitrogen) with recombinant baculovirus at an MOI of 5-10 and incubating at 28°C shaking at 110 RPM. After 72 hours, the cultures were clarified by two rounds of centrifugation at 2000g and 10,000g at 4 °C. The supernatant, containing secreted, soluble HA was concentrated and buffer exchanged into 1x PBS, pH 7.4. After metal affinity chromatography using Ni-NTA resin, HAs were modified and purified further as required for specific purposes (see following sections). At this stage, yields typically varied from 1-10mg/L, depending upon the HA isolate.

Biotinylation and Purification of HAs for Affinity Maturation and Binding Studies

After Ni-NTA purification, HAs with C-terminal biotinylation tags were concentrated down to ~2-5mg/mL total protein. The HAs were biotinylated by the addition of 25ug BirA enzyme/mg total protein, in a buffer of the following composition: 100mM Tris pH 8.0, 10mM ATP, 10mM MgOAc, 50uM biotin, with less than 50mM NaCl. The biotinylation reactions were incubated at 37°C for 1-2 hours. At this point, some HAs were digested with trypsin (New England Biolabs, 5mU trypsin per mg HA, 16 hours at 17°C) to generate the fusion competent HA1/HA2 form, while the majority were kept undigested as HA0. Biotinylated HAs were purified by size-exclusion chromatography, and concentrated down to ~5-20 mg/mL.

Expression and Purification of CR6261 Fab

Genes coding for the Fab region of the CR6261 heavy and light chains were synthesized (Mr. Gene), fused to the gp67 signal peptide and a C-terminal His tag by overlap PCR, and cloned into pFastBacDual (Invitrogen) for expression in baculovirus. Virus production methods, protein expression in High5 cells, harvesting, and Ni-NTA purification was essentially as described above for HA. CR6261 Fab was further purified by protein G affinity chromatography (elution in glycine buffer, pH 2.7); cation exchange chromatography (MonoS resin, sodium acetate, pH 5.0, with a linear gradient from 0-500mM NaCl); and gel filtration (10mM Tris, pH8.0, 150mM NaCl). The final yield was approximately 15mg/L.

Binder screening methodology

Designed binding proteins were tested for binding using yeast-surface display (S*15*). Yeast codon-optimized genes encoding designs were custom ordered from Genscript

(Piscataway, NJ) and subcloned between NdeI/XhoI sites in an in-house yeast display plasmid named pETCON™. pETCON™ is the original yeast display plasmid pCTCON(S*16*) with the following modifications: (a) a frameshift mutation in the CD20 encoding region; (b) a NdeI restriction site immediately downstream of the NheI site; and (c) a XhoI-Gly_2 spacer sequence immediately upstream of the BamHI restriction site. The full sequence is available upon request. Binding studies were done essentially as described (S*15*) using 1 μM of a biotinylated SC/1918/H1 HA1-2 ectodomain, except where noted otherwise. Secondary labels were anti-cmyc FITC (Miltenyi Biotec, Auburn, CA) to monitor design surface expression and streptavidin-phycoerythrin (Invitrogen, Carlsbad, CA) to monitor binding of the biotinylated antigen. Binding signal was quantified as the mean phycoerythrin fluorescence of the displaying population of cells using a 488 nm laser for excitation and a 575 nm band pass filter for emission (appropriately compensated) using either a Cytopeia inFlux Cell Sorter or an Accuri C6 flow cytometer.

The positive control for binding was CR6261 scFv Phe54Ala. The CR6261 scFv was constructed by a $(Gly_4Ser)_3$ linker joining the heavy to the light variable region using the DNA encoding CR6261 Fab (S*17*) as a template. The scFv was further amplified to include recombination sites for integration into pETCON between the NdeI/XhoI restriction sites. The Phe54Ala and all other point mutations were introduced by the method of Kunkel (S*18*).

Affinity Maturation

HB36 Round 1: First-generation libraries were constructed from the designed HB36 gene by error-prone PCR (epPCR) on the entire amino-acid coding segment or through single site-saturation mutagenesis at 22 out of the 27 residues that are modeled as being within 10Å from HA. In this and other cases, epPCR was done using a Stratagene GeneMorph™ II random mutagenesis kit (Agilent, CA) and site- saturation mutagenesis by the method of Kunkel (S*19*). The total library size was 3e5. We carried these libraries through 2 sorts of yeast display selection, with cells labeled at 50 nM HA1-2 for sort 1 and 10 nM for sort 2. Asp47X and Ala60Val/Thr mutations were recovered that improved affinity >10-fold. The best combination was used for the start of Round 2, and was HB36 Asp47Ser Ala60Val (HB36.3).

HB36 Round 2: Second-generation libraries were constructed from HB36.3 gene using epPCR at 2±1 mutations per gene. A total of 4 yeast display sorts were taken on a library size of

5.4e6. For the 1st sort, cells were labeled with 5 nM HA1-2 and gated to collect the top 5% of the population. For the second and third sorts, cells were labeled with 10 nM HA1-2 and then thoroughly washed with phosphate buffer saline with 1 mg/mL Fraction V bovine serum albumin (Sigma, St. Louis, Mo.). Cells were then incubated at 22°C with 1 μM of soluble HB36.3 for 40 min. A 4th sort was taken with an off-rate incubation of 60 min. All clones selected from this round included the mutation Asn64Lys.

HB80 Round 1: First-generation libraries were constructed from the designed HB80 gene by epPCR using a mutational load of 2±1 mutations per gene. A library of 1.6e6 transformants was subjected to selection using a labeling concentration of 1 μM HA1-2 and three total sorts. We recovered mutations Met26Val/Thr and Asn36Lys, each of which improved affinity >10-fold. A gene encoding a combination of these mutations HB80 Met26Thr Asn36Lys and a truncation after position 54 (named HB80.2) was the starting sequence for the next round of selection.

HB80 Round 2: Starting with the HB80.2 gene, an epPCR library with a mutational load of 2±1 mutations per gene was transformed into yeast, yielding 2e4 transformants. Cells were labeled with HA1-2 at 3 nM (sort 1), and 5 nM (sorts 2&3) and gated to collect the top 4-5% of cells. All clones selected had a Asp12Gly or a Ala24Ser mutation.

Protein-design expression and purification

Genes encoding the designs were subcloned (NdeI/XhoI) in a pET29b expression vector (EMD, Gibbstown, NJ) and transformed into *E. coli* Rosetta™ (DE3) chemically competent cells. Protein expression was induced using the autoinduction method of Studier (S*20*). After expression for 24 h at 18°C, cells were pelleted, resuspended into buffer HBS (20 mM Hepes, 150 mM NaCl pH 7.4), and sonicated to release cell lysate. Following clarification by centrifugation, supernatant was applied to a Nickel column for purification. Proteins were eluted by step elution at 250 mM imidazole in HBS. Size exclusion chromatography on a Superdex75 column was used as a finishing purification step into HBS buffer.

Surface-plasmon resonance (SPR) data and analysis

All SPR data were recorded on a Biacore model 2000 (Biacore, Uppsula, Sweden). A streptavidin (SA)-coated chip (Biacore) was coated with 200 or 400 response units (RU) of

biotinylated SC/1918/H1 HA1-2 ectodomain. A blank flow cell and a flow cell coated with 200 RU biotinylated lysozyme were used as negative controls. 150 μL of designed protein at a flowrate of 50 μL/min with a dissociation time of 900 s was used throughout. At least 8 varying concentrations of protein were used to determine kinetic and equilibrium fits. Binding kinetics were evaluated using a 1:1 Langmuir binding model. Proteins were in buffer HBS with 0.1%(v/v) P20 surfactant and 0.5 mg/mL carboxymethyl dextran sodium salt (Biacore, Uppsula, Sweden) to minimize nonspecific adsorption onto the SA chip. Scrubber-2 software (see web site cores.utah.edu/interaction/) was used to fit the data globally using standard double background subtracted values.

Binder cross-reactivity studies by biolayer interferometry

Binding of HB80.3 and CR6261 Fab to a panel of representative HA isolates was assayed by biolayer interferometry using an Octet Red™ instrument (ForteBio, Inc.). Biolayer interferometry is conceptually similar to surface plasmon resonance experiments in that a protein of interest is immobilized on a surface and then exposed to potential binding partners in solution. The binding of analytes to the immobilized protein changes the optical properties of the biosensors, leading to a shift in the wavelength of light reflected off the binding surface. This shift in wavelength can be measured in real-time, allowing the measurement of association and dissociation rates and, therefore, K_d. Biotinylated HAs, purified as described above, were used for these measurements. HAs at ~10-50 μg/mL in 1x kinetics buffer (1x PBS, pH 7.4, 0.01% BSA, and 0.002% Tween 20) were loaded onto streptavidin coated biosensors and incubated with varying concentrations of HB80.3 or CR6261 Fab in solution. All binding data were collected at 25°C. The experiments comprised 5 steps: 1. Baseline acquisition (60 s); 2. HA loading onto sensor (180 s); 3. Second baseline acquisition (180 s); 4. Association of the designed binder for the measurement of k_{on} (180 s); and 5. Dissociation of the binder for the measurement of k_{off} (180 s). 4-6 concentrations of each binder were used, with the highest concentration being 100 nM. Baseline and dissociation steps were carried out in buffer only. The sequences of all proteins used in this work are available in FASTA format as Table 10 below.

Expression and purification of HB36.3 for crystallization

Rosetta™ 2 (BL21/DE3) cells carrying the pET29a-HB36.3 construct were grown in shake flasks in low salt LB medium to an OD_{600} of ~0.7 at 37°C, then shifted to 18°C and induced by the addition of 1mM IPTG. Cultures were incubated overnight at 18°C for protein expression, then harvested by centrifugation (3000g, 10 minutes). The pellet from a 1L culture was resuspended in 50-100mL of lysis buffer (50mM Tris pH 8.0, 300mM NaCl, 10mM imidazole pH 8.0, with Roche EDTA-free protease inhibitor cocktail tablet) and the cells were lysed and homogenized by two passes through an EmulsiFlex™ C-3 cell disruptor (15kPSI). After clearing the lysates by centrifugation (25,000g, ~1 hour), the supernatant was incubated with NiNTA resin (Qiagen), washed with excess lysis buffer, and bound proteins were eluted (with 50mM Tris pH 8.0, 300mM NaCl, 250mM imidazole pH 8.0). The eluted material was buffer exchanged into 10mM Tris pH8.0, 50mM NaCl, loaded onto a MonoQ™ anion exchange column, and eluted with a linear gradient from 50-500mM NaCl. Peak fractions containing HB36.3 were pooled and subjected to gel filtration. HB36.3 eluted as an apparent dimer when loaded at high concentrations (~10mg/mL), but eluted as a monomer when loaded at lower concentrations (<1mg/mL), and the two forms were in rapid equilibrium. Fractions containing HB36.3 were pooled and concentrated to ~5mg/mL.

Isolation of HB36.3-SC1918/H1 HA complex for crystallization

Following Ni-NTA purification, SC1918 HA was digested with trypsin (New England Biolabs, 5mU trypsin per mg HA, 16 hours at 17°C) to produce uniformly cleaved (HA1/HA2), and to remove the trimerization domain and His-tag. After quenching the digests with 2mM PMSF, the digested material was purified by anion exchange chromatography (10mM Tris, pH 8.0, 50-1M NaCl) and size exclusion chromatography (10mM Tris, pH 8.0, 150mM NaCl). To prepare the HB36.3/SC1918 complex for crystallization, excess HB36.3 (approximately 5 HB36.3 molecules per HA trimer) was mixed with purified SC1918 HA in 10mM Tris pH 8.0, 150mM NaCl at ~2mg/mL. The mixtures were incubated overnight at 4°C to allow complex formation. Saturated complexes were then purified from unbound HB36.3 by gel filtration.

Crystallization and structure determination of the HB36.3-SC1918/H1 complex

Gel filtration fractions containing the HB36.3-SC1918/H1 HA complex were concentrated

to ~10mg/mL in 10mM Tris, pH 8.0 and 50mM NaCl. Initial crystallization trials were set up using the automated Rigaku Crystalmation™ robotic system at the Joint Center for Structural Genomics (web site JCSG.org). Several hits were obtained, with the most promising candidates grown in ~10% PEG8000 near pH 7. Optimization of these conditions resulted in diffraction quality crystals. The crystals used for data collection were grown by the sitting drop, vapor diffusion method with a reservoir solution (100uL) containing 10% PEG8000, 200mM magnesium chloride, and 100mM Tris pH 7.0. Drops consisting of 100nL protein + 100nL precipitant were set up at 4°C, and crystals appeared within 7-14 days. The resulting crystals were cryoprotected by soaking in well solution supplemented with increasing concentrations of ethylene glycol (5% steps, 5min/step), to a final concentration of 25%, then flash cooled and stored in liquid nitrogen until data collection.

Diffraction data for the HB36.3-SC1918/H1 complex were collected at the Advanced Photon Source (APS) General Medicine/Cancer Institutes-Collaborative Access Team (GM/CA-CAT) beamline 23ID-D at the Argonne National Laboratory. The data were indexed in R32, integrated using HKL2000 (HKL Research) and scaled using Xprep™ (Bruker). The structure was solved by molecular replacement to 3.10 Å resolution using Phaser™ (S*21*). An unpublished, 1.8 Å resolution structure of the 1918 HA was used as the initial search model and a single protomer was found in the asymmetric unit. Examination of the maps at this stage revealed clear positive electron density around the membrane distal end of HA consistent with the expected location and orientation of HB36.3. Attempts to place HB36.3 by molecular replacement using Phaser™ were unsuccessful (using various search models derived from PDB code 1U84). However, phasing using the HA only (~85% of the mass in the asymmetric unit) yielded maps with continuous density for HB36.3, including key side-chain features. This phasing model allowed HB36.3 to be fitted into the maps manually and unambiguously. Rigid-body and restrained refinement (including TLS refinement, with one group for HA1, one for HA2, and one for HB36.3) were carried out in Phenix™ (S*22*). Between rounds of refinement, the model was built and adjusted using Coot™ (S*23*). The insect cells used for protein expression produce fully glycosylated HA, and additional electron density was observed for glycans at all 5 predicted glycosylation sites (NX(S/T) motifs) on the HA. A total of 5 sugar residues were built at 2 of these sites (at the remaining three sites, density was too weak or ambiguous to allow accurate model building). The high redundancy of the relatively weak data

aided in obtaining relatively good quality electron density maps at this moderate resolution that were readily interpretable, particularly around the HB36.1-HA interface (see Fig. 4C), despite high apparent R_{sym} and B-values(S*24*).

Structural analyses

Hydrogen bonds and van der Waals contacts between HB36.3 and SC1918/H1 HA were calculated using HBPLUS™ (S*25*) and CONTACSYM™ (S*26*), respectively. Surface area burial was analyzed with Rosetta™ (S*27*). MacPyMol™ (DeLano Scientific) (S*28*) was used to render structure figures and for general manipulations. The final coordinates were validated using the JCSG quality control server (v2.7), which includes MolProbity™ (S*29*).

Protease susceptibility assay

Each reaction contained ~2.5 μg HA or ~5 μg binder-HA complex and 1% dodecyl-maltoside (to prevent aggregation of the post-fusion HA). Reactions were set up at room temperature and the pH was lowered by adding 100 mM buffer to all samples except controls. Sodium acetate was used for pH ranges 4.9 to 6.1, PIPES buffer for pH 6.2 to 7.4 and Tris for pH 7.5 and above. Reactions were thoroughly mixed, centrifuged at >12,000g for 30 seconds and allowed to incubate at 37 °C for one hour. After incubation, reactions were equilibrated to room temperature and the pH was neutralized by addition of 200 mM Tris, pH 8.5.
Trypsin was added to all samples except controls, at a final ratio of 1:25 for the SC1918/H1 reactions, and 1:50 for the Viet04/H5 reactions. SC1918/H1 and Viet04/H5 samples were digested overnight (18 hours) at 37 °C and 17 °C, respectively. Reactions were quenched by addition of non-reducing SDS buffer and were boiled for ~2 min. Samples were analyzed by SDS-PAGE.

Limitations of initial binding screen; other potential binders

One important component in the recovery of active binders from our design set is the choice of screening system. We chose the yeast surface- display assay as our screen because the system allows rapid testing of many designs, there was minimal non-specific adsorption of the biotinylated hemagglutinin (at 1 μM) on the yeast surface (low background), and the screen could be readily reconfigured to select for higher-affinity variants. While it has been reported

that binding dissociation constants are roughly equivalent between yeast display titrations and *in vitro* measurements (S*30*), we noted approximately 10-fold weaker affinity for *in vitro* SPR measurements as compared to the yeast surface display titrations. Although there may be many reasons for the discrepancies between measurements (*e.g.* non-specific lectin adsorption increasing the local HA concentration), we suspect that the major contribution to increased affinity on yeast is avidity effects between the trimeric ectodomain of hemagglutinin used for binding studies and the thousands of copies of designs displayed on the surface. Given the test concentration of 1 μM HA, we estimate that we were able to detect binding for designs that displayed on the surface of yeast with an *in vitro* $K_d < 25$ μM.

Another important parameter blocking the recovery of active designs is the dissociation rate of the HA-design complex. During affinity maturation of the HB36 & HB80 designs, we noted a marked increase in the mean phycoerythrin fluorescence (PE) signal at binding saturation for several variants, controlling for mean surface display of the design variants (data not shown). This increase in PE signal correlated with slower *in vitro* off-rates. Extrapolating the off-rate to the limit of binding detection by PE signal, we estimate that the yeast display system can detect binders with $k_{off} < 10\ s^{-1}$.

Thus our yeast display screen can recover from our design set all binders that surface display with an *in vitro* $K_d < 25$uM and a $k_{off} < 10s^{-1}$. Several designs showed weak binding activity in this screen; and include HB3, HB54, and HB78 (amino acid sequences are available in Table 9, below).

On the usefulness of *de novo* design in generating specific binders

As *de novo* design of protein interactions may find many uses, it is instructive to note the effort required to isolate the HA binders reported here. A number of technical advances coalesced to facilitate this research, including the availability of highly parallel computing, of yeast cell-surface display as a tool for fast screening and affinity maturation of binding proteins, the low cost of gene synthesis, and the ability to custom-order plasmids from commercial sources. For a typical *de novo* design goal, we estimate that a hundred thousand CPU hours would be sufficient to generate several dozen candidates for experimental testing. The yeast-display format used here removes the laborious steps needed for purifying each design and allows fast screening and affinity maturation.

While in this case two antibody-bound structures were available, the method made minimal use of information contained in these antibodies, with only a single hotspot residue in HB80 (the Tyr of HS3) coinciding with a residue on the antibodies. Only the structure of H1 HA was essential for the design process. The hemagglutinin target surface is very apolar, enabling the design of high-affinity interactions. It remains to be seen whether this methodology could be used to target more polar protein surfaces.

The importance of a diverse set of protein scaffolds for *de novo* design

The use of diverse protein folds was a crucial element in the success of the design method. Binding to the hydrophobic target site on HA is highly constrained due to flanking polar and charged loops and residues (Fig. 1). The backbones of both HB36 and HB80 are exquisitely suited to this site with helices that sequester their backbone polar groups from interacting with the apolar surface of HA, while the rest of the redesigned proteins form little if any interactions with the flanking HA regions (Figs. 1 & 2). The diversity of protein scaffolds available in the PDB has, therefore, been key to this design procedure. Nearly 40% of the proteins in the scaffold set were solved as part of the NIH NIGMS Protein Structure Initiative (PSI; web site is nigms.nih.gov/Initiatives/PSI/) and HB36 was derived from a PSI target protein of unknown function (APC36109 from *B. stearothermophilus*, PDB entry 1U84). While the utility to molecular biology of structures of relatively small, bacterial proteins of sometimes unknown function has been hotly debated by some (S*31-33*), we note that a previously unanticipated benefit of these structures is that they may open the road to the design of new protein functions.

Comparison of the designed proteins with post-fusion HA

Interestingly, the structure of post-fusion hemagglutinin (S*34*) reveals a helix bound to the hydrophobic region in the stem in a manner that is reminiscent of the main recognition helices observed in HB36 and HB80 although different in crucial details. The post-fusion structure shows significant rearrangement of the target epitope compared to the pre-fusion form, with the two loops that flank the hydrophobic surface moving away, providing unimpeded access to it. Against this surface, a helical segment from HA2 docks, burying the hydrophobic surface on the stem region. Although several hydrophobic chemical groups from this HA2 helix overlay on similar groups in the two designed binders , the angular orientation of the HA2 helix, its

length, and the identities of other residues preclude its use as a template from which to generate binders to the pre-fusion form. We nevertheless find this coarse similarity to be intriguingly suggestive of the phenomenon of structural mimicry(S*35*), whereby evolutionarily unrelated proteins present similar chemical groups for binding to certain target epitopes.

Table 9 FASTA sequences of active designs and design variants

>HB36.1 (Asp47Ser)
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAEAVLQAVYETESAFDLAMRIM
WIYAFAFNRPIPFPHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 270)

>HB36.2 (Ala60Val)
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAEAVLQAVYETEDAFDLAMRIM
WIYVFAFNRPIPFPHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 271)

>HB36.3 (Asp47Ser, Ala60Val)
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAEAVLQAVYETESAFDLAMRIM
WIYVFAFNRPIPFPHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 272)

>HB36.4 (Asp47Ser, Ala60Val, Asn64Lys)
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAEAVLQAVYETESAFDLAMRIM
WIYVFAFKRPIPFPHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 65)

>HB80
MASTRGSGRPWDFSENLAFELALAFMNKDTPDRWANVAQYVSGRTPEEVKKHYEILVE
DIKYIESGKVPFPNYRTTGGNMKTDEKRFRNLKIRLE (SEQ ID NO: 273)

>HB80 Met26Thr
MASTRGSGRPWDFSENLAFELALAFTNKDTPDRWANVAQYVSGRTPEEVKKHYEILVE
DIKYIESGKVPFPNYRTTGGNMKTDEKRFRNLKIRLE (SEQ ID NO: 180)

>HB80 Asn36Lys
MASTRGSGRPWDFSENLAFELALAFMNKDTPDRWAKVAQYVSGRTPEEVKKHYEILVE
DIKYIESGKVPFPNYRTTGGNMKTDEKRFRNLKIRLE (SEQ ID NO: 181)

>HB80.1 (Met26Thr, Asn36Lys)
MASTRGSGRPWDFSENLAFELALAFTNKDTPDRWAKVAQYVSGRTPEEVKKHYEILVE
DIKYIESGKVPFPNYRTTGGNMKTDEKRFRNLKIRLE (SEQ ID NO: 182)

>HB80.2 (Met26Thr, Asn36Lys, Delta54-95)
MASTRGSGRPWDFSENLAFELALAFTNKDTPDRWAKVAQYVSGRTPEEVKKHYE (SEQ ID NO: 183)

>HB80.3 (Asp12Gly, Ala24Ser, Met26Thr, Asn36Lys, Delta54-95)

MASTRGSGRPWGFSENLAFELALSFTNKDTPDRWAKVAQYVSGRTPEEVKKHYE (SEQ ID NO: 184)

>HB3
MADTLLILGDSLSAGYQMLAEFAWPFLLNKKWSKTSVVNASISGDTSQQGLARLPALL KQHQPRWVLVELGGNDGLEGFQPQQTEQTLRQILQDVKAANAEPLLMQIRPPANYGRR YNEAFSAIYPKLAKEFDVPLLPFFMEEVYLKPQWMQDDGIHPNYEAQPFIADWMAKQL QPLVNH (SEQ ID NO: 155)

>HB54
MAETKNFTDLVEATKWGNSLIKSAKYSSKDKMAIYNYTKNSSPINTPLRSANGDVNKLS ENIQEQVRQLDSTISKSVTPDSVYVYRLLNLDYLSSITGFTREDLHMLQQTNEGQYNSKL VLWLDFLMSNRIYRENGYSSTQLVSGAALAGRPIELKLELPKGTKAAYIDSKELTAYPG QQEVLLPRGTEYAVGTVELSKSSQKIIITAVVFKK (SEQ ID NO: 140)

>HB78
MFTGVIIKQGCLLKQGHTRKNWSVRKFILREDPAYLHYYYPLGYFSPLGAIHLRGCVVT SVESEENLFEIITADEVHYFLQAATPKERTEWIKAIQMASR (SEQ ID NO: 211)

Table 10 Sequences of HAs used in binding studies. The sequences listed below represent the full-length ORF as cloned in the baculovirus transfer vector. Most of the N-terminal signal peptide MVLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFA (SEQ ID NO: 212)) is presumably removed during secretion, leaving four non-native residues (ADPG) at the N-terminus of HA1. The C-terminal biotinylation site, trimerization domain, and His tag are retained on all.

>A/South Carolina/1/1918(H1N1)
MVLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADPGDTICIGYHANNSTDTV DTVLEKNVTVTHSVNLLEDSHNGKLCKLKGIAPLQLGKCNIAGWLLGNPECDLLLTASS WSYIVETSNSENGTCYPGDFIDYEELREQLSSVSSFEKFEIFPKTSSWPNHETTKGVTAAC SYAGASSFYRNLLWLTKKGSSYPKLSKSYVNNKGKEVLVLWGVHHPPTGTDQQSLYQ NADAYVSVGSSKYNRRFTPEIAARPKVRDQAGRMNYYWTLLEPGDTITFEATGNLIAP WYAFALNRGSGSGIITSDAPVHDCNTKCQTPHGAINSSLPFQNIHPVTIGECPKYVRSTKL RMATGLRNIPSIQSRGLFGAIAGFIEGGWTGMIDGWYGYHHQNEQGSGYAADQKSTQN AIDGITNKVNSVIEKMNTQFTAVGKEFNNLERRIENLNKKVDDGFLDIWTYNAELLVLL ENERTLDFHDSNVRNLYEKVKSQLKNNAKEIGNGCFEFYHKCDDACMESVRNGTYDYP KYSEESKLNREEIDGVSGGGGLNDIFEAQKIEWHERLVPRGSPGSGYIPEAPRDGQAYVR KDGEWVLLSTFLGHHHHHH (SEQ ID NO: 12)

>A/WSN/1933(H1N1)
MVLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADPGDTICIGYHANNSTDTV DTIFEKNVAVTHSVNLLEDRHNGKLCKLKGIAPLQLGKCNITGWLLGNPECDSLLPARS WSYIVETPNSENGACYPGDFIDYEELREQLSSVSSLERFEIFPKESSWPNHTFNGVTVSCS HRGKSSFYRNLLWLTKKGDSYPKLTNSYVNNKGKEVLVLWGVHHPSSSDEQQSLYSN GNAYVSVASSNYNRRFTPEIAARPKVKDQHGRMNYYWTLLEPGDTIIFEATGNLIAPWY AFALSRGFESGIITSNASMHECNTKCQTPQGSINSNLPFQNIHPVTIGECPKYVRSTKLRM

VTGLRNIPSIQYRGLFGAIAGFIEGGWTGMIDGWYGYHHQNEQGSGYAADQKSTQNAIN
GITNKVNSIIEKMNTQFTAVGKEFNNLEKRMENLNKKVDDGFLDIWTYNAELLVLLENE
RTLDFHDLNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCDNECMESVRNGTYDYPKY
SEESKLNREKIDGVSGGGGLNDIFEAQKIEWHERLVPRGSPGSGYIPEAPRDGQAYVRKD
GEWVLLSTFLGHHHHHH (SEQ ID NO: 13)

>A/AA/Marton/1943(H1N1)
MVLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADPGDTICIGYHANNSTDTV
DTVLEKNVTVTHSVNLLEDSHNGKLCRLKGIAPLQLGKCNIAGWILGNPECESLLSERS
WSYIVETPNSENGTCYPGDFIDYEELREQLSSVSSFERFEIFSKESSWPKHNTTRGVTAAC
SHAGKSSFYRNLLWLTEKDGSYPNLNNSYVNKKGKEVLVLWGVHHPSNIKDQQTLYQ
KENAYVSVVSSNYNRRFTPEIAERPKVRGQAGRMNYYWTLLKPGDTIMFEANGNLIAP
WYAFALSRGFGSGIITSNASMHECDTKCQTPQGAINSSLPFQNIHPVTIGECPKYVRSTKL
RMVTGLRNIPSIQSRGLFGAIAGFIEGGWTGMIDGWYGYHHQNEQGSGYAADQKSTQN
AINGITNKVNSVIEKMNTQFTAVGKEFNNLEKRMENLNKKVDDGFLDIWTYNAELLVL
LENERTLDFHDSNVKNLYEKVKNQLRNNAKEIGNGCFEFYHKCNNECMESVKNGTYD
YPKYSEESKLNREKIDSGGGGLNDIFEAQKIEWHERLVPRGSPGSGYIPEAPRDGQAYVR
KDGEWVLLSTFLGHHHHHH (SEQ ID NO: 14)

>A/USSR/90/1977(H1N1)
MVLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADPGDTICIGYHANNSTDTV
DTVLEKNVTVTHSVNLLEDSHNGKLCRLKGIAPLQLGKCNIAGWILGNPECESLFSKKS
WSYIAETPNSENGTCYPGYFADYEELREQLSSVSSFERFEIFPKERSWPKHNVTRGVTAS
CSHKGKSSFYRNLLWLTEKNGSYPNLSKSYVNNKEKEVLVLWGVHHPSNIEDQKTIYR
KENAYVSVVSSNYNRRFTPEIAERPKVRGQAGRINYYWTLLEPGDTIIFEANGNLIAPWH
AFALNRGFGSGIITSNASMDECDTKCQTPQGAINSSLPFQNIHPVTIGECPKYVRSTKLRM
VTGLRNIPSIQSRGLFGAIAGFIEGGWTGMIDGWYGYHHQNEQGSGYAADQKSTQNAIN
GITNKVNSVIEKMNTQFTAVGKEFNKLEKRMENLNKKVDDGFLDIWTYNAELLVLLEN
ERTLDFHDSNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCNNECMESVKNGTYDYPK
YSEESKLNREKIDSGGGGLNDIFEAQKIEWHERLVPRGSPGSGYIPEAPRDGQAYVRKDG
EWVLLSTFLGHHHHHH (SEQ ID NO: 43)

>A/Beijing/262/1995(H1N1)
MVLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADPGDTICIGYHANNSTDTV
DTVLEKNVTVTHSVNLLEDSHNGKLCRLKGIAPLQLGNCSVAGWILGNPECESLISKES
WSYIVETPNPENGTCYPGYFADYEELREQLSSVSSFERFEIFPKESSWPNHTVTGVTASCS
HNGKSSFYRNLLWLTEKNGLYPNLSNSYVNNKEKEVLVLWGVHHPSNIGVQRAIYHTE
NAYVSVVSSHYSRRFTPEIAKRPKVRGQEGRINYYWTLLEPGDTIIFEANGNLIAPWYAF
ALSRGFGSGIITSNAPMNECDAKCQTPQGAINSSLPFQNVHPVTIGECPKYVRSTKLRMV
TGLRNIPSIQSRGLFGAIAGFIEGGWTGMMDGWYGYHHQNEQGSGYAADQKSTQNAIN
GITNKVNSVIEKMNTQFTAVGKEFNKLERRMENLNKKVDDGFLDIWTYNAELLVLLEN
ERTLDFHDSNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCNNECMESVKNGTYDYPK
YSEESKLNREKIDSGGGGLNDIFEAQKIEWHERLVPRGSPGSGYIPEAPRDGQAYVRKDG
EWVLLSTFLGHHHHHH (SEQ ID NO: 54)

>A/Solomon Islands/3/2006(H1N1)

MVLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADPGDTICIGYHANNSTDTV
DTVLEKNVTVTHSVNLLEDSHNGKLCRLKGIAPLQLGNCSVAGWILGNPECELLISRES
WSYIVEKPNPENGTCYPGHFADYEELREQLSSVSSFERFEIFPKESSWPNHTTTGVSASCS
HNGESSFYKNLLWLTGKNGLYPNLSKSYANNKEKEVLVLWGVHHPPNIGDQRALYHK
ENAYVSVVSSHYSRKFTPEIAKRPKVRDQEGRINYYWTLLEPGDTIIFEANGNLIAPRYA
FALSRGFGSGIINSNAPMDECDAKCQTPQGAINSSLPFQNVHPVTIGECPKYVRSAKLRM
VTGLRNIPSIQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSGYAADQKSTQNAI
NGITNKVNSVIEKMNTQFTAVGKEFNKLERRMENLNKKVDDGFIDIWTYNAELLVLLE
NERTLDFHDSNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCNDECMESVKNGTYDYP
KYSEESKLNREKIDSGGGGLNDIFEAQKIEWHERLVPRGSPGSGYIPEAPRDGQAYVRKD
GEWVLLSTFLGHHHHHH (SEQ ID NO: 274)

>A/Japan/305/1957(H2N2)
MVLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADPGDQICIGYHANNSTEKV
DTILERNVTVTHAKDILEKTHNGKLCKLNGIPPLELGDCSIAGWLLGNPECDRLLSVPEW
SYIMEKENPRDGLCYPGSFNDYEELKHLLSSVKHFEKVKILPKDRWTQHTTTGGSRACA
VSGNPSFFRNMVWLTEKGSNYPVAKGSYNNTSGEQMLIIWGVHHPNDETEQRTLYQNV
GTYVSVGTSTLNKRSTPEIATRPKVNGQGGRMEFSWTLLDMWDTINFESTGNLIAPEYG
FKISKRGSSGIMKTEGTLENCETKCQTPLGAINTTLPFHNVHPLTIGECPKYVKSEKLVLA
TGLRNVPQIESRGLFGAIAGFIEGGWQGMVDGWYGYHHSNDQGSGYAADKESTQKAF
DGITNKVNSVIEKMNTQFEAVGKEFSNLERRLENLNKKMEDGFLDVWTYNAELLVLME
NERTLDFHDSNVKNLYDKVRMQLRDNVKELGNGCFEFYHKCDDECMNSVKNGTYDY
PKYEEESKLNRNEIKSGGGGLNDIFEAQKIEWHERLVPRGSPGSGYIPEAPRDGQAYVRK
DGEWVLLSTFLGHHHHHH (SEQ ID NO: 275)

>A/Hong Kong/1/1968(H3N2)
MVLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADPGATLCLGHHAVPNGTL
VKTITDDQIEVTNATELVQSSSTGKICNNPHRILDGIDCTLIDALLGDPHCDVFQNETWDL
FVERSKAFSNCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSNACKRGPGS
GFFSRLNWLTKSGSTYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTSLYVQASGRVT
VSTRRSQQTIIPNIGSRPWVRGLSSRISIYWTIVKPGDVLVINSNGNLIAPRGYFKMRTGKS
SIMRSDAPIDTCISECITPNGSIPNDKPFQNVNKITYGACPKYVKQNTLKLATGMRNVPEK
QTRGLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQINGKLNRVI
EKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTDSEMN
KLFEKTGRQLRENAEDMGNGCFKIYHKCDNACIESIRNGTYDHDVYRDEALNNRFQIKG
VSGGGGLNDIFEAQKIEWHERLVPRGSPGSGYIPEAPRDGQAYVRKDGEWVLLSTFLGH
HHHHH (SEQ ID NO: 276)

>A/duck/Czechoslovakia/1956 (H4N6)
MVLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADPGPVICMGHHAVANGTM
VKTLADDQVEVVTAQELVESQNLPELCPSPLRLVDGQTCDIINGALGSPGCDHLNGAEW
DVFIERPNAVDTCYPFDVPEYQSLRSILANNGKFEFIAEEFQWNTVKQNGKSGACKRAN
VNDFFNRLNWLVKSDGNAYPLQNLTKINNGDYARLYIWGVHHPSTDTEQTNLYKNNP
GRVTVSTKTSQTSVVPNIGSRPLVRGQSGRVSFYWTIVEPGDLIVFNTIGNLIAPRGHYKL
NNQKKSTILNTAIPIGSCVSKCHTDKGSLSTT (SEQ ID NO: 277)

References for Supplemental material

S1. J. Karanicolas *et al.*, *Mol. Cell*, in press (2011).

S2. J. J. Gray *et al.*, *J Mol Biol* **331**, 281 (2003).

S3. D. C. Ekiert *et al.*, *Science* **324**, 246 (2009).

S4. R. L. Dunbrack, Jr., M. Karplus, *Nat Struct Biol* **1**, 334 (1994).

S5. K. Henrick, J. M. Thornton, *Trends Biochem Sci* **23**, 358 (1998).

S6. D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H. J. Wolfson, *Nucleic Acids Res* **33**, W363 (2005).

S7. C. A. Smith, T. Kortemme, *J Mol Biol* **380**, 742 (2008).

S8. B. Kuhlman *et al.*, *Science* **302**, 1364 (2003).

S9. J. J. Havranek, D. Baker, *Protein Sci* **18**, 1293 (2009).

S10. T. Kortemme, D. Baker, *Proc. Natl. Acad. Sci. USA* **99**, 14116 (2002).

S11. M. C. Lawrence, P. M. Colman, *J Mol Biol* **234**, 946 (1993).

S12. S. Henikoff, J. G. Henikoff, *Proteins* **17**, 49 (1993).

S13. *Acta Crystallogr D Biol Crystallogr* **50**, 760 (1994).

S14. P. H. Brown, J. E. Cronan, M. Grotli, D. Beckett, *J Mol Biol* **337**, 857 (2004).

S15. G. Chao *et al.*, *Nat Protoc* **1**, 755 (2006).

S16. C. P. Graff, K. Chester, R. Begent, K. D. Wittrup, *Protein Eng Des Sel* **17**, 293 (2004).

S17. M. Throsby *et al.*, *PLoS One* **3**, e3942 (2008).

S18. L. M. Kunkel, A. P. Monaco, W. Middlesworth, H. D. Ochs, S. A. Latt, *Proc Natl Acad Sci U S A* **82**, 4778 (1985).

S19. T. A. Kunkel, *Proc Natl Acad Sci U S A* **82**, 488 (1985).

S20. F. W. Studier, *Protein Expr Purif* **41**, 207 (2005).

S21. A. J. McCoy *et al.*, *J Appl Crystallogr* **40**, 658 (2007).

S22. P. D. Adams *et al.*, *Acta Crystallogr D Biol Crystallogr* **66**, 213 (2010).

S23. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **66**, 486 (2010).

S24. Z. Dauter, *Acta Crystallogr D Biol Crystallogr* **55**, 1703 (1999).

S25. I. K. McDonald, J. M. Thornton, *J. Mol. Biol.* **238**, 777 (1994).

S26. S. Sheriff, W. A. Hendrickson, J. L. Smith, *J Mol Biol* **197**, 273 (1987).

S27. R. Das, D. Baker, *Annu Rev Biochem* **77**, 363 (2008).

S28. W. L. DeLano, *DeLano Scientific, Palo Alto, CA, USA*, (2002).
S29. V. B. Chen *et al.*, *Acta Crystallogr D Biol Crystallogr* **66**, 12 (2010).
S30. E. T. Boder, K. S. Midelfort, K. D. Wittrup, *Proc Natl Acad Sci U S A* **97**, 10701 (2000).
S31. T. A. Steitz, *Structure* **15**, 1523 (2007).
S32. S. K. Burley, A. Joachimiak, G. T. Montelione, I. A. Wilson, *Structure* **16**, 5 (2008).
S33. J. M. Chandonia, S. E. Brenner, *Science* **311**, 347 (2006).
S34. J. Chen, J. J. Skehel, D. C. Wiley, *Proc Natl Acad Sci U S A* **96**, 8967 (1999).
S35. C. E. Stebbins, J. E. Galan, *Nature* **412**, 701 (2001).

Example 2.

Yeast-displayed designs protect HA from undergoing pH-induced conformational change

SC1918/H1 HA was produced according to previous reports and was confirmed to be cleaved to HA1 & HA2 using denaturing gel electrophoresis. H1 HA was chemically biotinylated in PBS pH 7.4 at rt for 30 min using a 10-fold molar excess of sulfo-NHS-LC-biotin (Pierce), after which time the protein was desalted into 10 mM Tris, 150 mM NaCl, pH 8.0 using a desalting spin column (ThermoScientific) and stored at 4°C.

To determine whether H1 HA could undergo irreversible conformational changes in the absence of the protective effects of designs, 80 nM of H1 HA was incubated in a final volume of 100 uL at either buffer BBSF (20 mM BTP, 150 mM NaCl, 1 mg/mL Fraction V BSA, pH 7.4) or buffer pHBSF (100 mM sodium acetate, 150 mM NaCl, 1 mg/mL Fraction V BSA, pH 5.2) for 1 h at 37°C, after which the reactions were neutralized with 20 μL of 1 M Tris-HCl pH 8.0. Reaction mixtures were vortexed, spun at 20,000 x g for 5 min, and the supernatant at 10-fold dilution was used to label yeast cells displayed with either CR6261 scFv or affinity-matured designs. Cells were labeled for 30 min at 22°C in buffer, washed, and secondary labeled for 10 min on ice with anti-cmyc FITC (Miltenyi Biotec, Auburn, CA) and streptavidin-phycoerythrin (Invitrogen, Carlsbad, CA). After washing, cells were re-suspended in BBSF buffer and fluorescence of 20,000 cells was quantified using an Accuri C6 flow cytometer. Both CR6261 scFv and the designed binders target an epitope on HA that is absent in the post-fusion conformational change. H1 HA treated at pH 5.2 for 1 h had significantly lower fluorescence

relative to controls for all three surface-displayed HA binders, indicating that the H1 HA can undergo irreversible conformational change to the post-fusion state under these treatment conditions (data not shown).

To determine whether yeast-displayed designs can protect against H1 HA pH-induced conformational changes, 8 nM of H1 HA was used to label yeast cells displayed with either CR6261 scFv or affinity-matured designs. 2e6 cells were labeled for 30 min at 22°C in 1 mL BBSF buffer, washed once, and resuspended in either BBSF or pHBSF buffer and incubated at 37°C for 1-24 h. Periodically, samples were withdrawn in 100 μL volume and neutralized with 20 uL of 1 M Tris-HCl pH 8.0. Cells were pelleted, washed, and processed exactly as above. Sequential timepoints up to 24 h of this process were assessed. Notably, yeast cells displaying either the CR6261 scFv or the HB80.3 design show no significant difference in binding signal between the cells incubated in low pH or neutral pH buffer, showing that these designs most likely protect against the low-pH induced conformational change of H1 HA. Yeast cells displaying the HB36.4 design show no difference in binding signal between the low pH buffer and neutral pH buffer incubation until the 24 h timepoint, when a slight decrease in binding signal at the low pH incubation was seen.

Example 3. Profiling of the sequence-specific determinants of binding for designs using selections coupled to next generation DNA sequencing.

Methods

Library Creation

Single site saturation mutagenesis libraries for HB36.4 and HB80.3 were constructed from synthetic DNA by Genewhiz. Parental sequences are listed in Table 11 with mutagenic region highlighted in red. Yeast EBY100 cells were transformed with library DNA and linearized pETCON (Science, 2011) using established protocols, yielding 1.4e6 and 3.3e6 transformants for the HB36.4 & HB80.3 ssm libraries, respectively. After transformation, cells were grown overnight in SDCAA media in 30 mL cultures at 30°C, passaged once, and stored in 20 mM HEPES 150 mM NaCl pH 7.5, 20% (w/v) glycerol in 1e7 aliquots at -80°C.

Table 11. DNA sequences of the single site saturation mutagenesis libraries. Base in italics and enlarged font indicate start and end of design encoding sequence. Base pairs in bold font indicate region of single site saturation mutagenesis.

>HB36.4
GACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAGGCTAGTG
GTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTCGGCTAGC*CATAT*
*G*CACATGTCCAATGCTATGGATGGTCAACAATTGAACAGATTGTTATTGGAATGGA
TCGGTGCCTGGGAC**CCTTTTGGTTTGGGTAAAGATGCTTATGACGTCGAAGCCG**
AAGCTGTTTTACAAGCAGTATACGAAACTGAATCTGCATTTGATTTGGCCATGA
GAATTATGTGGATCTATGTTTTTGCCTTCAAGAGACCAATTCCTTTCCCACACG
CTCAAAAATTGGCAAGAAGATTATTGGAATTGAAGCAAGCTGCATCTTCACCTTT
ACCATTGGAA*CTCGAG*GGGGGGCGGATCCGAACAAAAGCTTATTTCTGAAGAGGA
CTTGTAATAGAGATCT (SEQ ID NO: 214)

>HB80.3
GACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAGGCTAGTG
GTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTCGGCTAGC*CATAT*
*G***GCTTCTACTAGAGGTTCTGGTAGACCTTGGGGTTTTTCCGAAAATTTGGCCTT**
CGAATTGGCTTTAAGTTTTACTAACAAAGATACACCAGACAGATGGGCTAAGGT
TGCACAATATGTATCTGGTAGAACACCTGAAGAAGTTAAAAAGCATTACGAA*C*
*TCGAG*GGGGGGCGGATCCGAACAAAAGCTTATTTCTGAAGAGGACTTGTAATAGAG
ATCT (SEQ ID NO: 215)

Yeast display selections

Cell aliquots were thawed on ice, centrifuged at 13,000 rpm for 30 s, resuspended in 1e7 cells per mL of SDCAA media, and grown at 30°C for 6 h. Cells were then centrifuged for 13,000 rpm and resuspended at 1e7 cells per mL SGCAA media and induced at 22°C between 16-24 h. Cells were labeled with either biotinylated Viet/2004/H5 HA or SC/1918/H1 HA, washed, secondary labeled with SAPE (Invitrogen) and anti-cmyc FITC (Miltenyi Biotech), and sorted by fluorescent gates as outlined in Table 12. Cells were recovered overnight at 2.5e5 collected cells per mL SDCAA media, whereupon at least 1e7 cells were spun down at 13,000 rpm for 1 min and stored as cell pellets at -80°C before library prep for deep sequencing.

Table 12 Summary of selection conditions for yeast populations deep sequenced.

Expt	Sample	Sort	Library	Labeling Condition	% Cells Collected	# Cells Collected

1	No Gate	1	HB36.4	--	--	2.5E+05
1	Display	1	HB36.4	--	100%	2.5E+05
1	H1 bind (stringent)	1	HB36.4	18 nM H1 HA	41%	2.5E+05
1	H1 bind	1	HB36.4	60nM H1 HA	45%	2.5E+05
1	H5 bind	1	HB36.4	36nM H5 HA	33%	1.5E+05
1	No Gate	2	HB36.4	--	--	2.5E+05
1	Display	2	HB36.4	--	100%	2.5E+05
1	H1 bind (stringent)	2	HB36.4	3.5 nM H1 HA	10%	1.6E+05
1	H1 bind	2	HB36.4	42 nM H1 HA	64%	2.5E+05
1	H5 bind (stringent)	2	HB36.4	6 nM H5 HA	6%	6.0E+04
2	No Gate	1	HB36.4	--	--	1.5E+05
2	H1 bind	1	HB36.4	4nM H1 HA	19%	1.5E+05
2	No Gate	2	HB36.4	--	--	1.5E+05
2	H1 off-rate	2	HB36.4	6nM H1, 120' off with HB80.3	3%	9.0E+04
2	No Gate	1	HB80.3	--	--	1.5E+05
2	H1 bind	1	HB80.3	4nM H1 HA	21%	1.5E+05
2	No Gate	2	HB80.3	--	--	1.5E+05
2	H1 off-rate	2	HB80.3	6nM H1 HA, 40' off with HB80.3	2%	6.0E+04
3	No Gate	1	HB36.4	--	--	5.0E+05
3	Display	1	HB36.4	--	100%	5.0E+05
3	Good Display	1	HB36.4	--	10%	5.0E+05
3	Weak Display	1	HB36.4	--	27%	5.0E+05
3	H5 bind	1	HB36.4	10 nM H5 HA	30%	5.0E+05
3	No Gate	2	HB36.4	--	--	5.0E+05
3	H5 off-rate	2	HB36.4	3 nM H5 HA, 20' off with HB36.4	3%	3.0E+05
3	No Gate	1	HB80.3	--	--	5.0E+05
3	Display	1	HB80.3	--	100%	5.0E+05
3	Good Display	1	HB80.3	--	9%	5.0E+05
3	Weak Display	1	HB80.3	--	20%	5.0E+05
3	H5 bind	1	HB80.3	10 nM H5 HA	37%	5.0E+05
3	No Gate	2	HB80.3	--	--	5.0E+05
3	H5 off-rate	2	HB80.4	3 nM H5 HA, 75' off with HB36.4	11%	5.0E+05

Library prep and sequencing

Between 1-4e7 yeast cells were resuspended in Solution I (Zymo Research yeast plasmid miniprep II kit) with 25 U zymolase and incubated at 37°C for 4 hrs. Cells were then freeze/thawed using a dry ice/ethanol bath and a 42°C incubator. Afterwards, plasmid was recovered using a zymo research yeast plasmid miniprep II kit (Zymo Research, Irvine, CA) into a final volume of 30 μL 10 mM Tris-HCl pH 8.0. Contaminant genomic DNA was processed (per 20 μL rxn) using 2 μL ExoI exonuclease (NEB), 1 μL lambda exonuclease (NEB), and 2 μL lambda buffer at 30°C for 90 min followed by heat inactivation of the enzymes at 80°C for 20 min. Plasmid DNA was separated from the reaction mixture using a Qiagen PCR cleanup kit (Qiagen). Next, 18 cycles of PCR (98°C 10 s, 68°C 30s, 72°C 10 s) using Phusion high fidelity polymerase (NEB, Waltham, MA) was used to amplify the template and add the Illumina adaptor sections. Primers used were sample-specific and are listed in Table 13. PCR reaction was purified using an Agencourt AMPure™ XP kit (Agencourt, Danvers, MA) according to the manufacturer's specifications. Samples were quantified using Qubit dsDNA HS kit (Invitrogen) for a final yield of 1-4 ng/uL. Samples were combined in an equimolar ratio; from this pool, 0.4 fmol of total DNA was loaded on 2 separate lanes and sequenced using a Genome Analyzer IIx (Illumina) with appropriate sequencing primers (Table 13).

Table 13. List of sequencing primers used.

Primer Name	Sequence	Use
PCR77_fwd	AATGATACGGCGACCACCGAGATCTACACcggctagccatatggcttct (SEQ ID NO: 216)	NG lib construction
PCR77_rev_BC1	CAAGCAGAAGACGGCATACGAGATCAAGGTCAgatccgcccccctcgag (SEQ ID NO: 217)	NG lib construction
PCR77_rev_BC10	CAAGCAGAAGACGGCATACGAGATACGTACTCgatccgcccccctcgag (SEQ ID NO: 218)	NG lib construction
PCR77_rev_BC11	CAAGCAGAAGACGGCATACGAGATCTTCTAAGgatccgcccccctcgag (SEQ ID NO: 219)	NG lib construction
PCR77_rev_BC12	CAAGCAGAAGACGGCATACGAGATACTATGACgatccgcccccctcgag (SEQ ID NO: 220)	NG lib construction
PCR77_rev	CAAGCAGAAGACGGCATACGAGATG	NG lib construction

_BC13	ACGTTAAgatccgcccccctcgag (SEQ ID NO: 221)	
PCR77_rev _BC14	CAAGCAGAAGACGGCATACGAGATA CAAGATAgatccgcccccctcgag (SEQ ID NO: 222)	NG lib construction
PCR77_rev _BC15	CAAGCAGAAGACGGCATACGAGATG ACTAAGAgatccgcccccctcgag (SEQ ID NO: 223)	NG lib construction
PCR77_rev _BC16	CAAGCAGAAGACGGCATACGAGATG TGTCTACgatccgcccccctcgag (SEQ ID NO: 224)	NG lib construction
PCR77_rev _BC17	CAAGCAGAAGACGGCATACGAGATT TCACTAGgatccgcccccctcgag (SEQ ID NO: 225)	NG lib construction
PCR77_rev _BC18	CAAGCAGAAGACGGCATACGAGATA ATCGGATgatccgcccccctcgag (SEQ ID NO: 226)	NG lib construction
PCR77_rev _BC19	CAAGCAGAAGACGGCATACGAGATA GTACCGAgatccgcccccctcgag (SEQ ID NO: 227)	NG lib construction
PCR77_rev _BC2	CAAGCAGAAGACGGCATACGAGATG CATAACTgatccgcccccctcgag (SEQ ID NO: 228)	NG lib construction
PCR77_rev _BC3	CAAGCAGAAGACGGCATACGAGATC TCTGATTgatccgcccccctcgag (SEQ ID NO: 229)	NG lib construction
PCR77_rev _BC30	CAAGCAGAAGACGGCATACGAGATG TAGCAGTgatccgcccccctcgag (SEQ ID NO: 230)	NG lib construction
PCR77_rev _BC31	CAAGCAGAAGACGGCATACGAGATG GATCATCgatccgcccccctcgag (SEQ ID NO: 231)	NG lib construction
PCR77_rev _BC32	CAAGCAGAAGACGGCATACGAGATG TGAACGTgatccgcccccctcgag (SEQ ID NO: 232)	NG lib construction
HA77_f1_fwd	Cggctagccatatggcttct (SEQ ID NO: 233)	NG sequencing
HA77_f1_rev	Gtgcaaccttagcccatctgtctggtg (SEQ ID NO: 234)	NG sequencing
HA77_f2_fwd	Ggccttcgaattggctttaagttttactaacaaagat (SEQ ID NO: 235)	NG sequencing
HA77_f2_rev	Gatccgcccccctcgag (SEQ ID NO: 236)	NG sequencing
HA77_index	Ctcgaggggggcggatc (SEQ ID NO: 237)	NG sequencing
PCR35_fw	AATGATACGGCGACCACCGAGATCT	NG lib construction

d	ACACgatcggtgcctgggac (SEQ ID NO: 238)	
PCR35_rev _BC20	CAAGCAGAAGACGGCATACGAGATT TGCCTCAcagcttgcttcaattccaataatc (SEQ ID NO: 239)	NG lib construction
PCR35_rev _BC21	CAAGCAGAAGACGGCATACGAGATT CGTTAGCcagcttgcttcaattccaataatc (SEQ ID NO: 240)	NG lib construction
PCR35_rev _BC22	CAAGCAGAAGACGGCATACGAGATT ATAGTTCcagcttgcttcaattccaataatc (SEQ ID NO: 241)	NG lib construction
PCR35_rev _BC23	CAAGCAGAAGACGGCATACGAGATT GGCGTATcagcttgcttcaattccaataatc (SEQ ID NO: 242)	NG lib construction
PCR35_rev _BC24	CAAGCAGAAGACGGCATACGAGATT GGACATGcagcttgcttcaattccaataatc (SEQ ID NO: 243)	NG lib construction
PCR35_rev _BC25	CAAGCAGAAGACGGCATACGAGATA GGTTGCTcagcttgcttcaattccaataatc (SEQ ID NO: 244)	NG lib construction
PCR35_rev _BC26	CAAGCAGAAGACGGCATACGAGATA TATGCTGcagcttgcttcaattccaataatc (SEQ ID NO: 245)	NG lib construction
PCR35_rev _BC27	CAAGCAGAAGACGGCATACGAGATG TACAGTGcagcttgcttcaattccaataatc (SEQ ID NO: 246)	NG lib construction
PCR35_rev _BC40	CAAGCAGAAGACGGCATACGAGATA ATCCTGCcagcttgcttcaattccaataatc (SEQ ID NO: 247)	NG lib construction
PCR35_rev _BC41	CAAGCAGAAGACGGCATACGAGATG TTATATCcagcttgcttcaattccaataatc (SEQ ID NO: 248)	NG lib construction
PCR35_rev _BC42	CAAGCAGAAGACGGCATACGAGATA CACACGTcagcttgcttcaattccaataatc (SEQ ID NO: 249)	NG lib construction
PCR35_rev _BC43	CAAGCAGAAGACGGCATACGAGATA TACGACTcagcttgcttcaattccaataatc (SEQ ID NO: 250)	NG lib construction
PCR35_rev _BC44	CAAGCAGAAGACGGCATACGAGATA TCTTCGTcagcttgcttcaattccaataatc (SEQ ID NO: 251)	NG lib construction
PCR35_rev _BC45	CAAGCAGAAGACGGCATACGAGATA CATGTATcagcttgcttcaattccaataatc (SEQ ID NO: 252)	NG lib construction
PCR35_rev _BC46	CAAGCAGAAGACGGCATACGAGATT CCACAGTcagcttgcttcaattccaataatc (SEQ ID NO: 253)	NG lib construction

PCR35_rev_BC47	CAAGCAGAAGACGGCATACGAGATCAGTCTGTcagcttgcttcaattccaataatc (SEQ ID NO: 254)	NG lib construction
HA35_f1_fwd	Gatcggtgcctgggac (SEQ ID NO: 255)	NG sequencing
HA35_f1_rev	Tcttgaaggcaaaaacatagatccacataattctcatgg (SEQ ID NO: 256)	NG sequencing
HA35_f2_fwd	Acaagcagtatacgaaactgaatctgcatttgatttgg (SEQ ID NO: 257)	NG sequencing
HA35_f2_rev	Cagcttgcttcaattccaataatc (SEQ ID NO: 258)	NG sequencing
HA35_index	Gattattggaattgaagcaagct (SEQ ID NO: 259)	NG sequencing
Up-GS-pCons	Ggacaatagctcgacgattgaaggtagatacccata (SEQ ID NO: 260)	Universal fwd primer
Down_Cmyc	Caagtcctcttcagaaataagcttttgttc (SEQ ID NO: 261)	Universal rev primer
HB80_front_rev	Tggtctaccggaacctctggtggatgc (SEQ ID NO: 262)	Elibrary construction
HB80_back_fwd	Actcctgaagaagtcaaaaagcattacgaa (SEQ ID NO: 263)	Elibrary construction
HB80_klenow	Ttcgtaatgctttttgacttcttc (SEQ ID NO: 264)	Elibrary construction
E80 ultramer	Gcatccaccagaggttccggtagaccatggrrgttcarsgaaaacvttrmgtttgaamttgctttgtmttttacgaataaggacacaccagatagatggrvgaaggttgcayrstatgtaarsggtagaactcctgaagaagtcaaaaagcattacgaa (SEQ ID NO: 265)	Elibrary construction
HB36_front_rev	Gtcataggcatctttacccaaacc (SEQ ID NO: 266)	Elibrary construction
HB36_back_fwd	Catgcccaaaagttggctaga (SEQ ID NO: 267)	Elibrary construction
HB36_klenow	Tctagccaacttttgggcatgt (SEQ ID NO: 268)	Elibrary construction
E36 ultramer	Ccttttggtttgggtaaagatgcctatgackwtgaagccgmtrvagttttamaggcagtatacgmgactramymtgcttttgacttggcaatgagaattmwktggatctatrwttttgcctwtaagagammgattcctttcvyacatgcccaaaagttggctaga (SEQ ID NO: 269)	Elibrary construction

Sequencing Analysis

Alignment and quality filtering of the sequencing data from raw Illumina reads were treated essentially as described previously. Each sequencing read was assigned to the correct

pool on the basis of a unique 8 bp barcode identifier (Table 13). All pools were treated identically in sequence analysis and quality filtration. Custom scripts were used to align all paired-end reads with both reads above an average Phred quality score equal or above 20. Paired-end reads were aligned using a global Needleman-Wunsch algorithm, reads without gaps were merged into a single sequence and differences between sequences resolved using the higher quality score for the read. Sequencing technical replicates of the naïve library indicate that the enumeration error for the library prep and sequencing falls under a poisson distribution; therefore, bootstrapping was used to estimate confidence intervals for error analysis. All error listed is at the 95% confidence interval.

Affinity Maturation and Specificity

Beneficial mutations predicted to result in higher affinity for SC1918/H1 HA were combined into a single library. The DNA library for each design was constructed from SOE PCR using a single oligo encoding the variable region. Primers and sequences are listed in Table 13, while the DNA sequence for the libraries are listed in Table 14. Libraries were transformed into yeast EBY100 and subjected to 4 sorts of varying stringency against biotinylated SC1918 H1/HA.

Table 14. DNA sequences of the affinity maturation libraries constructed from the information contained in the deep sequencing experiment.

>HB36.4_elibrary
GACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAGGCTAGTG
GTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTCGGCTAGCCATATG
CACATGTCCAATGCTATGGATGGTCAACAATTGAACAGATTGTTATTGGAATGGATC
GGTGCCTGGGAC
CCT TTT GGT TTG GGT AAA GAT GCT TAT GMT KWT GAA GCC GAA RVA GTT TTA
MAG GCA GTA TAC GMG ACT RAM YMT GCA TTT GAT TTG GCC ATG AGA ATT
MWK TGG ATC TAT RWT TTT GCC TWT AAG AGA MMG ATT CCT TTC VYA CAC
GCT CAA AAA TTG GCA AGA
AGATTATTGGAATTGAAGCAAGCTGCATCTTCACCTTTACCATTGGAACTCGAGGGG
GGCGGATCCGAACAAAAGCTTATTTCTGAAGAGGACTTGTAATAGAGATCT (SEQ ID NO: 201)

>HB80.3_elibrary
GACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAGGCTAGTG
GTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTCGGCTAGCCATATG

GCT TCT ACT AGA GGT TCT GGT AGA CCT TGG RRG TTT ARS GAA AAT VTT RMG TTC GAA MTT GCT TTA TMT TTT ACT AAC AAA GAT ACA CCA GAC AGA TGG RVG AAG GTT GCA YDS TAT GTA ARS GGT AGA ACA CCT GAA GAA GTT AAA AAG CAT TAC GAA
CTCGAGGGGGGCGGATCCGAACAAAAGCTTATTTCTGAAGAGGACTTGTAATAGAG ATCT (SEQ ID NO: 203)

For the HB36.4 epistatic library, no dominant lineage was converged after four sorts (Table 15). Promising constructs were subcloned (NdeI/XhoI) into the pET29b (Novagen) *E. coli* expression plasmid. For the HB80.4 epistatic library, clones after four sorts converged to two dominant lineages, each with at least 5 amino acid mutations from the starting HB80.3 sequence (Tables 16). Promising constructs were subcloned into a custom pET plasmid (NdeI/XhoI) with an N-terminal FLAG tag and a C-terminal His_6 tag and subjected to a solubility screen.

Table 15. FASTA sequences of selected constructs from the HB36.4 epistatic library after four sorts. All clones significantly outperform HB36.4 on yeast-surface display titrations.

>HB36.4_s4_E03
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDDEAAAVLQAVYETNHAFDLAMRIH WIYVFAFKRKIPFLHAQKLARRLLELKQAASSPLP (SEQ ID NO: 69)
>HB36.4_s4_E05
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAAAVLKAVYATNSAFDLAMRII WIYVFAYKRKIPFAHAQKLARRLLELKQAASSPLP (SEQ ID NO: 70)
>HB36.4_s4_E06
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDFEADKVLQAVYETNSAFDLAMRIN WIYVFAFKRPIPFVHAQKLARRLLELKQAASSPLP(SEQ ID NO: 71)
>HB36.4_s4_E07
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAAAVLKAVYETNSAFDLAMRIN WIYVFAFKRKIPFAHAQKLARRLLELKQAASSPLP(SEQ ID NO: 72)
>HB36.4_s4_E08
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEADKVLQAVYDTNSAFDLAMTIH WIYNFAFKRKIPFLHAPKLARRLLELKLAASSPLP (SEQ ID NO: 73)
>HB36.4_s4_E09
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDDEADRVLQAVYETNSAFDLAMRIN WIYVFAFKRTIPFAHAQKLARRLLELKQAASSPLP (SEQ ID NO: 74)
>HB36.4_s4_E10
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDYEADKVLQAVYETNSAFDLAMRIH WIYIFAFKRPIPFVHAQKLARRLLELKQAASSPLP (SEQ ID NO: 75)
>HB36.4_s4_E11

MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEADAVLKAVYETNSAFDLAMRIH
WIYNFAFKRKIPFVHAQKLARRLLELKQAASSPLP (SEQ ID NO: 76)
>HB36.4_s4_E12
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDDEADKVLQAVYATNSAFDLAMRIH
WIYNFAYKRTIPFVHAQKLARRLLELKQAASSPLP (SEQ ID NO: 77)
>HB36.4_s4_E13
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDDEAARVLKAVYATDSAFDLAMRIH
WIYNFAFKRKIPFLHAQKLARRLLELKQAASSPLP (SEQ ID NO: 78)
>HB36.4_s4_E14
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEADKVLQAVYATNSAFDLAMRIH
WIYIFAFKRTIPFIHAQKLARRLLELKQAASSPLP (SEQ ID NO: 79)
>HB36.4_s4_E17
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDYEADEVLKAVYATNSAFDLAMRIH
WIYNFAFKRKIPFTHAQKLARRLLELKQAASSPLP (SEQ ID NO: 80)
>HB36.4_s4_E18
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAAKVLQAVYETNSAFDLAMKIH
WIYNFAFKRTIPFVHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 81)
>HB36.4_s4_E19
MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEADKVLQAVYATNSAFDLAMKIH
WIYIFAFKRTIPFIHAQKLARRLLELKQAASSPLP (SEQ ID NO: 82)

Table 16. FASTA sequences of selected constructs from the HB80.3 epistatic library after four or five sorts. All clones significantly outperform HB80.3 on yeast-surface display titrations.

>HB80.3_s4_E81
MASTRGSGRPWRFSENVAFEIALSFTNKDTPDRWKKVARYVRGRTPEEVKKHYE (SEQ ID NO: 187)

>HB80.3_s4_E82
MASTRGSGRPWKFSENVAFEIALSFTNKDTPDRWAKVARYVRGRTPEEVKKHYE (SEQ ID NO: 188)
>HB80.3_s4_E83
MASTRGSGRPWGFRENIAFEIALYFTNKDTPDRWRKVARYVKGRTPEEVKKHYE (SEQ ID NO: 189)
>HB80.3_s4_E84
MASTRGSGRPWRFSENVAFEIALSFTNKDTPDRWRKVARYVRGRTPEEVKKHYE (SEQ ID NO: 190)
>HB80.3_s4_E85
MASTRGSGRPWGFSENIAFELALYFTNKDTPDRWGKVARYVRGRTPEEVKKHYE (SEQ ID NO: 191)
>HB80.3_s4_E86
MASTRGSGRPWKFSENVAFELALYFTNKDTPDRWKKVARYVKGRTPEEVKKHYE (SEQ ID NO: 192)
>HB80.3_s4_E87
MASTRGSGRPWKFSENIAFELALYFTNKDTPDRWKKVARYVKGRTPEEVKKHYE (SEQ ID NO: 193)

>HB80.3_s4_E88
MASTRGSGRPWKFKENLEFEIALSFTNKDTPDRWKKVAYYVRGRTPEEVKKHYE (SEQ ID NO: 194)
>HB80.3_s4_E89
MASTRGSGRPWRFSENVAFEIALSFTNKDTPDRWRKVARYVRGRTPEEVKKHYE (SEQ ID NO: 190)
>HB80.3_s4_E90
MASTRGSGRPWKFSENVAFELALYFTNKDTPDRWTKVARYVKGRTPEEVKKHYE (SEQ ID NO: 196)
>HB80.3_s4_E91
MASTRGSGRPWKFSENVAFELALYFTNKDTPDRWTKVARYVKGRTPEEVKKHYE (SEQ ID NO: 196)
>HB80.3_s4_E92
MASTRGSGRPWKFSENVAFEIALSFTNKDTPDRWRKVARYVRGRTPEEVKKHYE (SEQ ID NO: 198)
>HB80.3_s4_E93
MASTRGSGRPWKFSENVAFELALYFTNKDTPDRWGKVAQYVRGRTPEEVKKHYE (SEQ ID NO: 199)
>HB80.3_s4_E94
ASTRGSGRPWKFSENVAFELALYFTNKDTPDRWAKVARYVKGRTPEEVKKHYE (SEQ ID NO: 200)
>HB80.3_s4_E95
MASTRGSGRPWKFSENVAFELALYFTNKDTPDRWTKVARYVKGRTPEEVKKHYE (SEQ ID NO: 196)
>HB80.3_s4_E96
MASTRGSGRPWKFSENVAFEIALSFTNKDTPDRWRKVAYYVRGRTPEEVKKHYE (SEQ ID NO: 202)
>HB80.3_s4_E97
MASTRGSGRPWRFSENVAFEIALSFTNKDTPDRWRKVARYVRGRTPEEVKKHYE (SEQ ID NO: 190)
>HB80.3_s4_E98
MASTRGSGRPWRFSENVAFEIALSFTNKDTPDRWAKVARYVRGRTPEEVKKHYE (SEQ ID NO: 204)
>HB80.3_s4_E99
MASTRGSGRPWKFSENLAFELALYFTNKDTPDRWAKVAYYVKGRTPEEVKKHYE (SEQ ID NO: 205)
>HB80.3_s4_E100
MASTRGSGRPWRFSENVAFEIALSFTNKDTPDRWKKVARYVKGRTPEEVKKHYE (SEQ ID NO: 206)
>HB80.3_s5_E01
MASTKGSGKPWKFSENVAFEIALSFTNKDTPDRWRKVARYVRGKTPEEVKKHYE (SEQ ID NO: 207)
>HB80.3_s5_E04
MASTRGSGRPWKFSENVAFEIALSFTNKDTPDRWRKVARYVRGRTPEEVKKHYE (SEQ ID NO: 198)
>HB80.3_02

MASTRGSGRPWKFSENIAFEIALSFTNKDTPDRWKKVAQYVKGRTPEEVKKHYE (SEQ ID NO: 209)
>HB80.3_16
MASTRGSGRPWKFSENIAFEIALSFTNKDTPDRWKKVAQYVKGRTPEEVKKHYE (SEQ ID NO: 209)

Solubility Screening

HB80.3 clones selected from the affinity maturation library were screened by solubility in an *E. coli* expression system using a dot-blot assay. Cells were grown from colonies in deep well plates overnight, and diluted 25-fold into deep well plates at 37°C for 3 h, followed by IPTG induction (1 mM) for 4 h at 37°C. Following induction, cells were separated from spent media by centrifugation at 3,000 x g for 15 min at 4°C and stored as pellets overnight at -20°C. The next morning, plates were thawed on ice for at least 15 min and 200 uL binding buffer (200 mM HEPES, 150 mM NaCl, pH 7.5) was added to each well. The plate was sonicated using the Ultrasonic Processor 96-well sonicator for 3 min at 70% pulsing power and lysate centrifuged for 4000 rpm for 30 min at 4°C. Supernatant at 100-fold dilution was transferred to a dot blot manifold Minifold™ I (Whatman) and dried onto nitrocellulose membrane for 5 min. The membrane was then labeled with an anti-FLAG HRP conjugated mouse antibody (Sigma, St. Louis, MO) and visualized with DAB substrate (Pierce).

Table 17 provides per position allowable substitutions on an HB36.4 scaffold.

HB36.4:Central helix recognition motif from Serine 47- Phenylalanine 63 (SAFDLAMRIMWIYVFAF (SEQ ID NO: 7)); Also Phe 69 outside of that recognition motif (MSNAMDGQQLNRLLLEWIGAWDPFGLGKDAYDVEAEAVLQAVYETESAFDLAMRIM WIYVFAFKRPIPFPHAQKLARRLLELKQAASSPLPLE (SEQ ID NO: 65))

(2) Allowable positions were determined from yeast display selections of HB36.4 variants to SC1918/H1 HA coupled to deep sequencing (see attached for further details). The threshold was no more than 80% depletion in the frequency of a given mutant in the selection library after two selection sorts by FACS. Positions listed in bold font indicate positions that make contact with the HA surface.

Table 17 Allowable substitutions on an HB36.4 scaffold

Position	**HB36.4 Residue**	**Allowable**
47 R1	Ser	ala,phe,his,lys,met,asn,gln,thr,val,tyr,
48 R2	Ala	All Amino Acids

49	**Phe**	**Phe**
50 R3	**Asp**	**Ala,Glu,Gly,Asn,Pro,Ser,Tyr**
51 R4	Leu	Phe
52 R5	Ala	All Amino Acids
53 R6	**Met**	**Phe,His,Ile,Leu,Gln,Thr**
54 R7	Arg	gly,lys,gln,thr
55 R8	Ile	asn,gln,val,trp
56 R9	**Met**	**Gly,Ile,Lys,Leu,Asn,Arg,Ser,Thr,Val,Tyr, His**
57 R10	**Trp**	**Phe**
58 R11	Ile	phe,ser,thr,val
59 R12	Tyr	cys,asp,phe,his,asn,ser
60 R13	**Val**	**Ala,Phe,Ile,Leu,Asn,Gln,Thr,Tyr**
61 R14	**Phe**	**Glu,Leu**
62 R15	Ala	gly,lys,arg,ser
63 R16	Phe	cys,his,lys,leu,met,asn,gln,arg,thr,val,trp,tyr
69 R17	**Phe**	**Tyr**

The table below shows where single point mutants from HB36.4 (SAFDLAMRIMWIYVFAF (SEQ ID NO: 7)) are shown to result in increased binding affinity.

Table 18 HB36.4 point mutations resulting in increased binding affinity

Position	**HB36.4 Residue**	**Increased Affinity**
47 R1	Ser	His
54 R7	Arg	Lys
56 R9	Met	His, Asn, Tyr
60 R13	Val	Phe, Leu, Thr, Asn
63 R16	Phe	Tyr

The table below provides per position allowable substitutions on an HB80.3 scaffold.

(1 Central helix recognition motif from Phenylalanine 13- Phenylalanine 25; Also Tyrosine 40 that is outside of that recognition motif).

(MASTRGSGRPWG**FSENLAFELALSF**TNKDTPDRWAKVAQ**Y**VSGRTPEEVKKHYE (SEQ ID NO: 184))

Allowable positions were determined from yeast display selections of HB80.3 variants to SC1918/H1 HA coupled to deep sequencing (see attached for further details). The threshold was no more than 80% depletion in the frequency of a given mutant in the selection library after two selection sorts by FACS. Positions listed in bold font indicate positions that make contact with the HA surface.

Table 19 Allowable substitutions on an HB80.3 scaffold

Position	**HB80.3 Residue**	**Allowable**

13 R1	**Phe**	**Val**
14 R2	**Ser**	**Ala,Phe,Gly,Ile,Lys,Leu,Met,Asn,Pro,Gln,Arg,Thr,Val**
15 R3	Glu	Asp
16 R4	Asn	His,Ile,Lys,Leu,Met,Arg,Ser,Thr
17 R5	**Leu**	**Phe,Ile,Met,Asn,Gln,Val**
18 R6	**Ala**	**Asp,Lys,Met,Asn,Gln,Arg,Val**
19 R7	Phe	Asp,Asn,Tyr
20 R8	Glu	Ala,Asp,Gly,His,Lys,Leu,Met,Asn,Gln,Arg,Ser,Thr,Val,Trp
21 R9	**Leu**	**Phe,Ile,Met,Val**
22	Ala	Ala
23 R10	Leu	Ile,Met,Tyr
24 R11	Ser	Ala,Gly,Tyr
25	**Phe**	**Phe**
39 **R12**	Gln	Tyr,Phe,Met,Arg,Lys,Gly
40 **R13**	Tyr	Asp,Met,Asn,Ser
42 **R14**	Ser	Arg,Lys

The table below shows where single point mutants from HB80.3 are shown to result in increased binding affinity.

Table 20 HB80.3 point mutations resulting in increased binding affinity

Position	**HB80.3 Residue**	**Increased Affinity**
14 R2	Ser	Ala, Gly,Ile,Lys,Arg,Thr,Val
17 R5	Leu	Ile,Val
18 R6	Ala	Lys,Arg
20 R8	Glu	Ser
21 R9	Leu	Ile
24 R11	Ser	Tyr

We claim:

1. A polypeptide comprising an amino acid sequence according to general formula I R1-R2-Phe-R3-R4-R5-R6-R7-R8-R9-R10-R11-R12-R13-R14-R15-R16 (SEQ ID NO: 1), wherein

R1 is selected from the group consisting of Ser, Ala, Phe, His, Lys, Met, Asn, Gln, Thr, Val, Tyr, and Asp;

R2 can be any amino acid;

R3 is selected from the group consisting of Asp, Ala, Glu, Gly, Asn, Pro, Ser, and Tyr;

R4 is selected from the group consisting of Leu and Phe;

R5 can be any amino acid;

R6 is selected from the group consisting of Met, Phe, His, Ile, Leu, Gln, and Thr;

R7 is selected from the group consisting of Arg, Gly, Lys, Gln, and Thr;

R8 is selected from the group consisting of Ile, Asn, Gln, Val, and Trp;

R9 is selected from the group consisting of Met, Gly, Ile, Lys, Leu, Asn, Arg, Ser, Thr, Val, His, and Tyr;

R10 is selected from the group consisting of Trp and Phe;

R11 is selected from the group consisting of Ile, Phe, Ser, Thr, and Val;

R12 is selected from the group consisting of Tyr, Cys, Asp, Phe, His, Asn, and Ser;

R13 is selected from the group consisting of Val, Ala, Phe, Ile, Leu, Asn, Gln, Thr, and Tyr;

R14 is selected from the group consisting of Phe, Glu, and Leu;

R15 is selected from the group consisting of Ala, Gly, Lys, Arg, and Ser; and

R16 is selected from the group consisting of Phe, Cys, His, Lys, Leu, Met, Asn, Gln, Arg, Thr, Val, Trp, and Tyr.

2. The polypeptide of claim 1, wherein general formula I is R1-R2-Phe-R3-R4-R5-R6-R7-R8-R9-R10-R11-R12-R13-R14-R15-R16-X1-R17 (SEQ ID NO: 2), wherein

X1 is 4-8 amino acids in length, wherein each position can be any amino acid; and

R17 is Phe or Tyr.

3. A polypeptide comprising an amino acid sequence according to general formula II R1-R2-R3-R4-R5-R6-R7-R8-R9-Ala-R10-R11-Phe (SEQ ID NO: 83), wherein

R1 is selected from the group consisting of Phe and Val;

R2 is selected from the group consisting of Ser, Ala, Phe, Gly, Ile, Lys, Leu, Met, Asn, Pro, Gln, Arg, Thr, and Val;

R3 is selected from the group consisting of Glu, and Asp;

R4 is selected from the group consisting of Asn, His, Ile, Lys, Leu, Met, Arg, Ser, and Thr;

R5 is selected from the group consisting of Leu, Phe, Ile, Met, Asn, Gln, and Val;

R6 is selected from the group consisting of Ala, Asp, Lys, Met, Asn, Gln, Arg, Glu, and Val;

R7 is selected from the group consisting of Phe, Asp, Asn, and Tyr;

R8 is selected from the group consisting of Glu, Ala, Asp, Gly, His, Lys, Leu, Met, Asn, Gln, Arg, Ser, Thr, Val, and Trp;

R9 is selected from the group consisting of Leu, Phe, Ile, Met, and Val;

R10 is selected from the group consisting of Leu, Ile, Met, and Tyr; and

R11 is selected from the group consisting of Ser, Ala, Gly, and Tyr;

4. The polypeptide of claim 3, wherein general formula II is R1-R2-R3-R4-R5-R6-R7-R8-R9-Ala-R10-R11-Phe-**X1-R12-R13-X2-R14 (SEQ ID NO: 84)**, wherein

X1 is 5-15 amino acids in length, wherein each position can be any amino acid;

R12 is selected from the group consisting of Gln, Tyr, Phe, Met, Arg, Lys, and Gly;

R13 is selected from the group consisting of Tyr, Asp, Met, Asn, and Ser;

X2 is any amino acid; and

R14 is selected from the group consisting of Ser, Arg, and Lys.

5. A polypeptide comprising an amino acid sequence selected from the group consisting of

(d) MADTLLILGDSLSAGYQMLAEFAWPFLLNKKWSKTSVVNASISGDTSQQGL ARLPALLKQHQPRWVLVELGGNDGLEGFQPQQTEQTLRQILQDVKAANAEP LLMQIRPPANYGRRYNEAFSAIYPKLAKEFDVPLLPFFMEEVYLKPQWMQDD GIHPNYEAQPFIADWMAKQLQPLVNH (SEQ ID NO: 155);

(e) MAETKNFTDLVEATKWGNSLIKSAKYSSKDKMAIYNYTKNSSPINTPLRSAN GDVNKLSENIQEQVRQLDSTISKSVTPDSVYVYRLLNLDYLSSITGFTREDLH MLQQTNEGQYNSKLVLWLDFLMSNRIYRENGYSSTQLVSGAALAGRPIELKL ELPKGTKAAYIDSKELTAYPGQQEVLLPRGTEYAVGTVELSKSSQKIIITAVVF KK (SEQ ID NO: 140); and

(f) MFTGVIIKQGCLLKQGHTRKNWSVRKFILREDPAYLHYYYPLGYFSPLGAIHL RGCVVTSVESEENLFEIITADEVHYFLQAATPKERTEWIKAIQMASR (SEQ ID NO: 211).

6. The polypeptide according to any one of claims 1-5, wherein the polypeptide comprises a detectable tag.
7. An isolated nucleic acid encoding the polypeptide of any one of claims 1-6.
8. A recombinant expression vector comprising the nucleic acid of claim 7, operatively linked to a suitable control sequence.
9. A recombinant host cell comprising the recombinant expression vector of claim 8.
10. An antibody that selectively binds to the polypeptide of any one of claims 1-6.
11. A pharmaceutical composition, comprising one or more polypeptides according to any one of claims 1-6 and a pharmaceutically acceptable carrier.
12. A method for treating and/or limiting an influenza infection, comprising administering to a subject in need thereof a therapeutically effective amount of one or more polypeptides of the invention, salts thereof, conjugates thereof, or pharmaceutical compositions thereof, to treat and/or limit the influenza infection.
13. A method for diagnosing an influenza infection, or monitoring progression of an influenza infection, comprising
    (a) contacting a biological sample from a subject suspected of having an influenza infection with a diagnostically effective amount of one or more polypeptides of the invention under conditions suitable for binding of the polypeptide to a viral HA protein present in the sample; and
    (b) detecting polypeptide-viral HA binding complexes,

where the presence of such binding complexes indicates that the subject has an influenza infection, or provides a measure progression of an influenza infection.

14. A method for identifying candidate influenza vaccines, comprising
    (a) contacting test compounds with a polypeptide of the present invention under conditions suitable for polypeptide binding;
    (b) removing unbound test compounds; and
    (c) identifying those test compounds that bind to the polypeptide of the invention, wherein such test compounds are candidate influenza vaccines.

15. A method for identifying candidate compounds for treating, limiting, and/or diagnosing influenza infection, comprising
    (a) contacting an influenza HA protein with (i) test compounds and (ii) a polypeptide of the present invention, under conditions suitable for binding of the HA protein to the polypeptide of the present invention; and
    (b) identifying those test compounds that outcompete the polypeptide for binding to the HA protein, wherein such test compounds are candidate compounds for treating, limiting, and/or diagnosing influenza infection.