Supporting Information

Greenbaum et al. 10.1073/pnas.0911580106

SI Methods

Querying for Epitopes with Human Relevance in the IEDB. Human-relevant influenza A epitopes were retrieved from the IEDB and stored in a local database. For T-cell epitopes with nested sequences, the optimal epitope(s) were chosen manually by comparing the assay descriptions associated with each epitope in the IEDB to identify the peptide with the most frequent responses or to choose multiple peptides if the epitopes were restricted by different MHC molecules. For linear B-cell epitopes with nested sequences, the shortest sequence $> = 7$ residues was chosen as the representative. For discontinuous epitopes no removal of nested sequences was performed.

The following query parameters (in boldface type) were used to extract epitopes with human relevance from the IEDB:

Epitope source organism = influenza A. We consider epitopes that were derived from any influenza A strain.

B-cell assay or (T-cell assay AND (host = human OR restricting MHC = human). We limit ourselves to epitopes characterized in experiments that demonstrate the presence of adaptive immune receptors (antibody/BCR or TCR) that recognize the epitope. For TCR recognition, the epitopes are presented by MHC molecules. As these molecules are highly divergent in different species, we limit T-cell epitopes to those identified in the presence of human MHC molecules.

Positive Measurement. The outcome of the experiment has to be reported as a positive response against the epitope. We make no attempt to enforce a common set of criteria for defining immunogenicity and protective efficacy, because widely divergent methodologies were used by different laboratories to measure immune responses. Rather, we record, for each epitope the specific assay category and conditions used, and conform to the criteria for defining positive and negative measurements as reported by the authors themselves in each published article.

Exclude Antigen = Epitope AND Immunogen = Epitope. We excluded epitopes that were defined solely by their use as both immunogen (to induce the responses) and as antigen (to measure the response). In such experiments, it is not possible to evaluate the relevance of the induced response for antiviral immunity, as the structural context of the epitope is missing. We considered only those epitopes shown to be recognized by Abs or T Cell Receptors (TCRs) in the context of the whole influenza virus or proteins.

Exclude small sequences. Linear sequences of <7 residues and discontinuous sequences <3 residues. These constitute partial epitope sequences and will be discarded, as their conservation cannot be accurately determined.

Querying for Epitopes Conserved in Circulating H1N1. The epitopes in the human-relevant set described above were searched against H3N2 and H1N1 sequences from circulating strains between 1988 and 2008 epitopes that were 100% conserved in 30% or more of the strains of any given year were kept in the set. This 30% frequency requirement was implemented to prevent the possible inclusion of rare H3N2 and H1N1 isolates. The conservation of discontinuous sequences was determined in a two-step process. First, a regular expression was created to represent the discontinuous epitope residues and the spacing between them. This was matched against the query sequence. If a match was found, the corresponding linear sequence that covers the epitope was extracted from the epitope source sequence and BLASTed against the query sequence. If the match occurred in the homologous region it was considered conserved.

Query Paramters for Influenza Sequences. The following query parameters were set:

Virus Species = Influenza A AND (Subtype = H1N1 OR Subtype = H3N2).
Host = Human
Year FROM 1979 TO 2008
Whole genome only

Swine flu sequences were obtained from two sources. The NCBI the NCBI Influenza Viral resource was queried for:

Virus Species = Influenza A AND Subtype = H1N1
Host = Human
Year 2009

Homology Modeling, Image Generation, and ASA Calculation. The homology model for Fig. 1 was generated using the SWISS-MODEL homology-modeling server (1). All six individual chains of the 1RUY PDB structure were modeled separately against the representative swine influenza HA sequence (ACP41934.1). The 1RUY structure was selected for modeling by BLASTing the ACP41934.1 sequence against the PDB (2), using the "Advanced Search" option on the PDB website. This structure is the best BLAST hit that does not include a synthetic construct. The chains were consolidated into one PDB file using UCSF Chimera (3). Coordinates of the trimer-fused hemagglutinin (PDB ID: 1HA0) were obtained from the PQS service (4) and the epitope was mapped onto the protein using the IEDB homology mapping tool (5). The images were produced using the WebLabViewer software (Accelrys Inc.). Residue surface accessibilities were calculated using NACCESS (6).

1. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31:3381–3385.
2. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242.
3. Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
4. Henrick K, Thornton JM (1998) PQS: A protein quaternary structure file server. *Trends Biochem Sci* 23:358–361.
5. Zhang Q, et al. (2008) Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res* 36(Web Server issue):W513–W518.
6. Hubbard S, JM T (1993) NACCESS (Department of Biochemistry and Molecular Biology, University College London).

Table S1. Epitope sequences conserved in swine-origin H1N1 influenza virus (S-OIV)

IEDB ID	Protein	Sequence	Epitope category
20836	HA	GLFGAIAGF	B-cell linear
37043	M1	LKTRPILSPLTKGILGFVFTLTVPSERGLQRRRFVQNALNGNGD	B-cell linear
97650	M2	SLLTEVET	B-cell linear
15381	NP	FDERRNKYLEEHPSAGKDPKKTGGPI	B-cell linear
17539	NP	FQTAAQR	B-cell linear
45359	NP	NPGNAEIEDLIFLAR	B-cell linear
67436	NP	TYQRTRALV	B-cell linear
97236	PB1	DAVATTHSWIPKRNRSIL	B-cell linear
20837	HA	GLFGAIAGFI	T-cell class I
97183	M1	AGKNTDLEALMEWLKTR	T-cell class I
27350	M1	ILSPLTKGIL	T-cell class I
20356	M1	GILGFVFTLTV	T-cell class I
54953	M1	RMVLASTTAK	T-cell class I
4349	M1	ASCMGLIY	T-cell class I
33844	M1	KTRPILSPLTK	T-cell class I
97506	M1	MSLLTEVETYVLSII	T-cell class I
51249	M1	QKRMGVQMQRFK	T-cell class I
59323	M1	SLLTEVETYVL	T-cell class I
28309	M1	IRHENRMVL	T-cell class I
53918	M1	RGLQRRRFVQNALNGNG	T-cell class I
20354	M1	GILGFVFTL	T-cell class I
58567	M1	SIIPSGPLK	T-cell class I
62486	NA	SWPDGAELPF	T-cell class I
97772	NP	YERMCNILKG	T-cell class I
97173	NP	AEIEDLIFLA	T-cell class I
60867	NP	SRYWAIRTR	T-cell class I
97614	NP	RMVLSAFDER	T-cell class I
53890	NP	RGINDRNFW	T-cell class I
55738	NP	RRSGAAGAAVK	T-cell class I
35590	NP	LELRSRYWAI	T-cell class I
35589	NP	LELRSRYWA	T-cell class I
97298	NP	FEDLRVSSF	T-cell class I
27283	NP	ILRGSVAHK	T-cell class I
27126	NP	ILKGKFQTA	T-cell class I
97178	NP	AFDERRNKYLEEHPSAGK	T-cell class I
7136	NP	CTELKLSDY	T-cell class I
13263	NP	ELRSRYWAI	T-cell class I
97583	NP	QLVWMACHSAA	T-cell class I
19312	NS1	GEISPLPSL	T-cell class I
97405	NS2	ITFMQALQLL	T-cell class I
97503	PA	MRRNYFTAEVSHCRATEY	T-cell class I
17119	PA	FMYSDFHFI	T-cell class I
1166	PA	AESRKLLLI	T-cell class I
62180	PA	SVKEKDMTK	T-cell class I
97781	PB1	YRRPVGISSMVEAMVSRA	T-cell class I
42143	PB1	MMMGMFNML	T-cell class I
21574	PB1	GPATAQMAL	T-cell class I
4177	PB1	ARLGKGYMF	T-cell class I
32289	PB1	KMARLGKGY	T-cell class I
10514	PB1	DTVNRTHQY	T-cell class I
70898	PB1	VSDGGPNLY	T-cell class I
97299	PB1	FEFTSFFY	T-cell class I
97693	PB1	TLARSICEK	T-cell class I
97314	PB1	FVANFSMEL	T-cell class I
65880	PB1	TQIQTRRSF	T-cell class I
16681	PB1	FLKDVMESM	T-cell class I
6174	PB1	CEKLEQSGL	T-cell class I
17780	PB1	FSMELPSFGV	T-cell class I
97309	PB1	FNMLSTVLGV	T-cell class I
97682	PB1	TFPYTGDPPYSHGTGTGY	T-cell class I
63635	PB1	TFEFTSFFY	T-cell class I
97779	PB2	YMLERELVRKTRFLPVA	T-cell class I
95905	HA	TGMVDGWYGYHHQNEQGS	T-cell class II
96007	HA	WTYNAELLVLLENERTLD	T-cell class II

IEDB ID	Protein	Sequence	Epitope category
95623	HA	NKVNSVIEKMNTQFTAVG	T-cell class II
97482	M1	LTKGILGFVFTLTVPSER	T-cell class II
97613	M1	RMVLASTTAKAMEQM	T-cell class II
97403	M1	IRHENRMVLASTTAKAM	T-cell class II
97740	M1	VLASTTAKAMEQMAGSSEQA	T-cell class II
67496	M1	TYVLSIIPSGPLKAEIAQRL	T-cell class II
21087	M1	GLQRRRFVQNALNGNGDPNN	T-cell class II
65112	M1	TLTVPSERGLQRRRFVQNAL	T-cell class II
37217	M1	LLENLQAYQKRMGVQMQRFK	T-cell class II
97418	M1	KGILGFVFTLTVPSE	T-cell class II
65389	M1	TNPLIRHENRMVLASTTAKA	T-cell class II
97730	M1	VFTLTVPSERGLQRRRFV	T-cell class II
2754	M1	ALMEWLKTRPILSPLTKGIL	T-cell class II
1579	M1	AGKNTDLEALMEWLKTRPIL	T-cell class II
97280	M1	ERGLQRRRFVQNALNGNG	T-cell class II
32182	NP	KLSTRGVQIASNEN	T-cell class II
97448	NP	LILRGSVAHKSCLPACVY	T-cell class II
45297	NP	NPAHKSQLVWMACHSAAFED	T-cell class II
97411	NP	KATNPIVPSFDMSNEGSY	T-cell class II
97487	NP	LVWMACHSAAFEDLR	T-cell class II
14070	NP	ERRNKYLEEHPSAGKDPKKT	T-cell class II
41793	NP	MIWHSNLNDATYQRTRALVR	T-cell class II
97637	NP	SFDMSNEGSYFFGDNA	T-cell class II
49220	NP	PRMCSLMQGSTLPRRSGAAG	T-cell class II
70712	NP	VRESRNPGNAEIEDLIFLARS	T-cell class II
97306	NP	FLARSALILRGSVAHK	T-cell class II
97416	NP	KFQTAAQRAMMDQVRESR	T-cell class II
9745	NP	DPRMCSLMQGSTLP	T-cell class II
97609	NP	RMCNILKGKFQTAAQRAM	T-cell class II
67439	NP	TYQRTRALVRTGMDP	T-cell class II
36692	NP	LIRMIKRGINDRNFWRGENG	T-cell class II
7655	NP	DATYQRTRALVRTGMDPRMC	T-cell class II
38689	NP	LPRRSGAAGAAVKG	T-cell class II
97361	NP	GQISVQPTFSVQRNLPF	T-cell class II
97269	NP	ELIRMIKRGINDRNFWR	T-cell class II
36863	NP	LKGKFQTAAQRAMMDQVRES	T-cell class II
10014	NS1	DRLRRDQKS	T-cell class II
97623	PA	RSKFLLMDALKLSIE	T-cell class II
97701	PB1	TNTETGAPQLNPIDGPL	T-cell class II
97354	PB1	GMFNMLSTVLGVSILNL	T-cell class II
97655	PB1	SPGMMMGMFNMLSTV	T-cell class II
97392	PB1	IFENSCLETMEVVQQTRV	T-cell class II
97389	PB1	HRGDTQIQTRRSFELKKL	T-cell class II
97559	PB1	PQLNPIDGPLPEDNEPSGY	T-cell class II
97228	PB1	CKLVGINMSKKKSYINK	T-cell class II
97611	PB1	RMFLAMITYITRNQPEWF	T-cell class II
97237	PB1	DCVLEAMAFLEESHPGIF	T-cell class II
97747	PB1	VNRTHQYSEKGKWTTNTE	T-cell class II
97610	PB1	RMFLAMITYITRNQP	T-cell class II
97353	PB1	GLPVGGNEKKAKLANVVR	T-cell class II
97489	PB1	MAFLEESHPGIFENS	T-cell class II
97713	PB1	TVLGVSILNLGQKKYTK	T-cell class II
97704	PB1	TQGRQTYDWTLNRNQPAA	T-cell class II
97501	PB1	MMGMFNMLSTVLGVS	T-cell class II
97519	PB2	NFVNRANQRLNPMHQLLR	T-cell class II

Table S2. Peripheral blood mononuclear cell (PBMC) donor information

ID	Sex	Age
ID104	M	35
ID110	M	41
ID111	M	41
ID113	M	27
ID114	M	26
ID115	M	27
ID117	M	28
ID130	M	36
ID132	F	62
ID140	M	34
BC-4	N/A	N/A
BC-11	N/A	N/A
BC-12	N/A	N/A
BC-15	N/A	N/A
BC-17	N/A	N/A
BC-19	N/A	N/A
BC-24	N/A	N/A
BC-34	N/A	N/A
BC-37	N/A	N/A
BC-40	N/A	N/A

F, female; M, male; N/A, not available. Sex and age of the PBMC donors is listed where available.

Table S3. All H1N1 isolates used in the analysis

Database	Isolate	N
NCBI	Influenza A virus [A/Auckland/1/2009(H1N1)]	3
NCBI	Influenza A virus [A/Auckland/2/2009(H1N1)]	2
NCBI	Influenza A virus [A/Auckland/3/2009(H1N1)]	3
NCBI	Influenza A virus [A/California/04/2009(H1N1)]	18*
NCBI	Influenza A virus [A/California/05/2009(H1N1)]	8
NCBI	Influenza A virus [A/California/06/2009(H1N1)]	10*
NCBI	Influenza A virus [A/California/07/2009(H1N1)]	23*
NCBI	Influenza A virus [A/California/08/2009(H1N1)]	13
NCBI	Influenza A virus [A/California/09/2009(H1N1)]	4
NCBI	Influenza A virus [A/California/10/2009(H1N1)]	3
NCBI	Influenza A virus [A/Denmark/513/2009(H1N1)]	6
NCBI	Influenza A virus [A/Italy/01/2009(H1N1)]	1
NCBI	Influenza A virus [A/Italy/02/2009(H1N1)]	2
NCBI	Influenza A virus [A/Kansas/03/2009(H1N1)]	1
NCBI	Influenza A virus [A/Netherlands/602/2009(H1N1)]	2
NCBI	Influenza A virus [A/New York/06/2009(H1N1)]	6
NCBI	Influenza A virus [A/New York/10/2009(H1N1)]	9
NCBI	Influenza A virus [A/New York/11/2009(H1N1)]	8
NCBI	Influenza A virus [A/New York/12/2009(H1N1)]	5
NCBI	Influenza A virus [A/New York/15/2009(H1N1)]	7
NCBI	Influenza A virus [A/New York/18/2009(H1N1)]	10*
NCBI	Influenza A virus [A/New York/19/2009(H1N1)]	10
NCBI	Influenza A virus [A/New York/20/2009(H1N1)]	2
NCBI	Influenza A virus [A/New York/23/2009(H1N1)]	5
NCBI	Influenza A virus [A/New York/31/2009(H1N1)]	6
NCBI	Influenza A virus [A/Ohio/07/2009(H1N1)]	14
NCBI	Influenza A virus [A/Regensburg/Germany/01/2009(H1N1)]	3
NCBI	Influenza A virus [A/Switzerland/01/2009(H1N1)]	1
NCBI	Influenza A virus [A/Texas/04/2009(H1N1)]	20*
NCBI	Influenza A virus [A/Texas/05/2009(H1N1)]	15
NCBI	Influenza A virus [A/Texas/06/2009(H1N1)]	7
NCBI	Influenza A virus [A/Toronto/3141/2009(H1N1)]	1
NCBI	Influenza A virus [A/Toronto/3145/2009(H1N1)]	1
NCBI	Influenza A virus [A/Toronto/3146/2009(H1N1)]	1
NCBI	Influenza A virus [A/Toronto/3170/2009(H1N1)]	1
NCBI	Influenza A virus [A/Toronto/3178/2009(H1N1)]	1
NCBI	Influenza A virus [A/Toronto/3181/2009(H1N1)]	1
NCBI	Influenza A virus [A/Toronto/3184/2009(H1N1)]	1
GISAID	Influenza A/Mexico/4108/2009	14
GISAID	Influenza A/Mexico/4115/2009	7
GISAID	Influenza A/Mexico/4482/2009	6
GISAID	Influenza A/Mexico/4486/2009	12
GISAID	Influenza A/Mexico/4603/2009	7
GISAID	Influenza A/Mexico/4604/2009	6

NCBI, National Center for Biotechnology Information.

*Indicates that there were sequences from each of the genome segments, excluding PB1-F2.