ARTICLE

Preparing for an avian flu pandemic: Method for designing possible subunit vaccines against highly pathogenic avian influenza viruses

Cecilia P. Mikita[1], Eduardo A. Padlan*[2]

[1]Walter Reed National Military Medical Center, Department of Medicine, Allergy/Immunology/Immunizations Service 4954 North Palmer Road, Bethesda, MD 20889-5600, USA
[2]4006 Simms Drive, Kensington, MD 20895-1336, USA

A method is described for designing possible subunit vaccines that focus the immune response to the cleavage site of the hemagglutinin molecule of highly pathogenic zoonotic avian influenza viruses. Cleavage of the hemagglutinin is critical to the entry of the virus into target cells. Sequence and three-dimensional structural data are analyzed to map the putative epitopes of antibodies against the hemagglutinin molecule, with particular emphasis on the epitope that includes the cleavage-site of the molecule. The chemical reactivity of the residues outside the cleavage-site epitope is reduced by judicious amino acid replacements so that the resulting molecule would be expected to elicit an immune response predominantly directed at the cleavage site, thereby preventing entry of the virus into target cells. The method is described with emphasis on highly pathogenic zoonotic H5N1 avian influenza viruses; the method could be used also on other avian influenza subtypes.

INTRODUCTION

Pandemics with highly pathogenic influenza A (HPAI) viruses have caused devastating morbidity and mortality in humans over centuries. The Spanish flu pandemic in 1918 was caused by an H1N1 influenza A virus, which killed more than 50 million people worldwide. Three additional influenza A pandemics have occurred in 1957 (H2N2), 1968 (H3N2), and 2009 (H1N1 "swine"), with significant human morbidity and mortality.

A member of the *Orthomyxoviridae* family, influenza viruses are single-stranded RNA viruses with four main species: A, B, C, and D. Influenza A species include all avian influenza viruses,

*Corresponding author
Email Address: eduardo.padlan@gmail.com
Date received: July 16, 2022
Date revised: October 26, 2022
Date accepted: October 26, 2022

KEYWORDS

highly pathogenic zoonotic avian influenza viruses, subunit vaccines, hemagglutinin, de-Antigenization

with zoonotic infection linked to direct or indirect contact with infected live or dead poultry or contaminated environments. Influenza A viruses are classified into subtypes by the hemagglutinin (HA) and neuraminidase (NA) surface proteins with 18 different HA proteins currently identified. Due to the rapid mutation rate of influenza viruses, HPAI viruses are capable of zoonotic transmission due to novel surface proteins that can spread within an immunologically naive population. Migratory birds have been implicated in the transcontinental spread of HPAI viruses, from wild to domestic avian populations to human transmission, substantiating the global threat of future avian influenza pandemics.

Initially identified in 1997, H5N1 influenza A has been implicated in zoonotic infections and human fatalities. HPAI H5N1 has continued to mutate, episodically affecting birds and humans on multiple continents, raising concerns for a future pandemic. Clinically, influenza A infection commonly presents with high fever and cough and progresses to lower respiratory tract symptoms of difficulty breathing. Additionally, H5N1 infection is also associated with gastrointestinal symptoms including nausea, vomiting, and diarrhea. Complications of influenza infection include pneumonia, acute respiratory distress syndrome, secondary bacterial infections, septic shock, or death.

The World Health Organization Global Influenza Strategy 2019-2030 has three main goals, to reduce the burden of seasonal influenza, minimize the risk of zoonotic influenza, and mitigate the impact of pandemic influenza. One of the main strategic objectives is to promote research and innovation for improved and novel diagnostics, vaccines, and treatments against influenza. In this regard, we will demonstrate our design on H5N1 avian influenza viruses, in particular those that had been shown to be highly pathogenic and zoonotic to humans. We have used the method previously in our design of possible subunit vaccines against seasonal influenza (Mikita and Padlan 2021) that focuses the immune response to the cleavage site of the hemagglutinin of the virus.

To focus the immune response to the cleavage site region of hemagglutinin, the method described by Padlan (2010) is used. Briefly, the method enhances the antigenicity of a chosen region by reducing the antigenicity of all the other regions. The method of Padlan (2010) assigns chemical reactivities (Sandberg et al. 1998, De Genst et al. 2002) to each amino acid in the structure; the chemical reactivity represents the contribution of an amino acid to the antigenicity of a particular epitope. The chemical reactivities used here are the ZZ3 values from Sandberg et al. (1998) (see also De Genst et al. 2002): the value of -3.5 is assigned to Arg, 1.93 to Asp, -0.11 to Glu, -2.49 to Lys, 1.15 to Ser, 1.04 to Asn, -1.44 to Gln, 0.36 to Gly, 1.84 to Pro, -1.12 to Thr, 0.60 to Ala, 0.26 to His, 3.71 to Cys, 0.47 to Met, -1.54 to Val, -1.71 to Ile, -1.49 to Leu, 0.43 to Tyr, 1.06 to Phe, and 0.59 to Trp. To reduce the contribution of an amino acid to antigenicity, referred to here as de-Antigenization, the amino acid is replaced by another of lower reactivity while ensuring the preservation of the overall structure of the antigen. The structure preservation is achieved by the replacement rules of Padlan (2008) shown in Table 1.

The hemagglutinin on the surface of influenza virus is a trimer of identical subunits and is used by the virus to gain entry into the target cell. First, the hemagglutinin is cleaved by proteolysis into two parts: HA1, the head region, and HA2, the stem. The cleavage results in a major structural change in the molecule (see, for example: Wiley et al. 1981). It is the cleaved hemagglutinin that allows entry into the target cell; therefore, prevention of this cleavage would be an ideal strategy to prevent viral infection. Antibody binding to the cleavage site of the hemagglutinin could effectively interfere with the proteolysis. The cleavage-site region of the hemagglutinin is the focus of the vaccine design being proposed.

MATERIALS AND METHODS

Structural and Sequence data analyzed

The method proposed for reducing the antigenicity of antibody epitopes requires the three-dimensional structure of the protein of interest. A number of hemagglutinin structures are available in the Protein Data Bank (PDB). However, the hemagglutinin structures currently available in PDB are not immediately usable in view of the fact that those hemagglutinins are already cleaved into HA1 and HA2 presumably as a consequence of isolation, purification, or other processes used during preparation or crystallization. Model building, using SWISS-MODEL (Peitsch 1995, Waterhouse et al. 2018) (https://swissmodel.expasy.org), was done to produce the three-dimensional structures that de-Antigenization requires.

Sequences for the hemagglutinin of known highly pathogenic avian influenza (HPAI) viruses were obtained from GenBank (https://www.ncbi.nlm.nih.gov/genbank). Only complete sequences in which all the residues had been identified were included in our collection. We limited our analysis to the sequences of the hemagglutinins of HPAI H5N1 that have been shown to be zoonotic to humans. Representatative sequences of HPAI H5N1 are shown in Table 2. The cleavage of hemagglutinin is at the site with the arginine at position 330 of the molecule (following the numbering in Table 2). A listing of the cleavage-site regions of the HPAI H5N1 sequences that we analyzed (Table 3) shows different gaps in the cleavage-site region.

One hemagglutinin structure available in the PDB (Entry 1HA0) is uncleaved; it is an H3N2 hemagglutinin in which the usual arginine at the cleavage site had been replaced by a glutamine (Wilson et al. 1981), thus preventing the proteolysis. One can only assume that the structure of the cleavage-site region in 1HA0 is identical to that in the native structure despite the arginine-to-glutamine replacement. It is also known that the result of the cleavage is a major change in the structure of the hemagglutinin (see, for example: Wiley et al. 1981). What, if any, structural changes had already occurred during the isolation, purification and crystallization of the cleaved hemagglutinin structures in PDB can only be guessed. The HPAI H5N1 hemagglutinin structures that we studied were modeled prior to further analysis using 1HA0 as the modeling template.

Calculation of antigenicities

A method has been proposed that could direct the antibody response to a chosen region of a protein antigen (Padlan 2010). That method is used here to design subunit vaccines against HPAI H5N1. The method reduces the antigenicity of putative antibody epitopes by judicious replacements of the residues in the epitope with residues of lower reactivity (see the replacement rules in Table 1). By reducing the antigenicity of all the putative antibody epitopes except the one that contains the cleavage-site, the antibody response could be expected to be directed predominantly against the cleavage-site.

The method developed earlier (Padlan 2008) was used for the characterization of putative antibody epitopes. The method defines a putative epitope as consisting of all the residues within a certain radius of a Cα (alpha-carbon) position of the molecule and assigns an antigenicity value for that epitope based on the summed reactivities of those residues weighted by their exposure to solvent. The exposures to solvent are calculated by

TABLE 1: The amino acid parameters used in the calculation of antigenicities and the replacement suggestions

Amino acid	zz1	zz2	zz3	zz4	zz5	SDGly	Propensities				If in Helix Sheet Coil Turn Change to:			
							Helix	Sheet	Coil	Turn				
Ala	0.24	-2.32	0.60	-0.14	1.30	60.0	0.00	0.47	-0.26154	0.83	-	-	-	-
Arg	3.52	2.50	-3.50	1.99	-0.17	125.0	0.21	0.35	-0.17659	0.82	Ala	Thr	Ala	Ala
Asn	3.05	1.62	1.04	-1.15	1.61	80.0	0.65	0.40	0.22989	1.44	Ala	Thr	Ser	Gly
Asp	3.98	0.93	1.93	-2.46	0.75	94.0	0.69	0.72	0.22763	1.41	Ala	Thr	Ser	Gly
Cys	0.84	-1.67	3.71	0.18	-2.65	159.0	0.68	0.25	-0.015152	1.08	-	-	-	-
Gln	1.75	0.50	-1.44	-1.34	0.66	87.0	0.39	0.34	-0.187677	0.94	Ala	Thr	Ala	Thr
Glu	3.11	0.26	-0.11	-3.04	-0.25	98.0	0.40	0.35	-0.20469	1.01	Ala	Thr	Ala	Thr
Gly	2.05	-4.06	0.36	-0.82	-0.38	0.0	1.00	--	0.43323	1.48	-	-	-	-
His	2.47	1.95	0.26	3.90	0.09	98.0	0.56	0.37	-0.0012174	1.07	Ala	Thr	Thr	Thr
Ile	-3.89	-1.73	-1.71	-0.84	0.26	135.0	0.41	0.10	-0.42224	0.59	-	-	-	-
Leu	-4.28	-1.30	-1.49	-0.72	0.84	138.0	0.21	0.32	-0.33793	0.66	-	-	-	-
Lys	2.29	0.89	-2.49	1.49	0.31	127.0	0.26	0.34	-0.100092	1.01	Ala	Thr	Thr	Thr
Met	-2.85	-0.22	0.47	1.94	-0.98	127.0	0.24	0.26	-0.22590	0.57	-	-	-	-
Phe	-4.22	1.94	1.06	0.54	-0.62	153.0	0.54	0.13	-0.22557	0.89	Ala	Thr	Ala	Ala
Pro	-1.66	0.27	1.84	0.70	2.00	42.0	3.01	--	0.55232	1.38	-	-	-	-
Ser	2.39	-1.07	1.15	-1.39	0.67	56.0	0.50	0.30	0.14288	1.15	-	-	-	-
Thr	0.75	-2.18	-1.12	-1.46	-0.40	59.0	0.66	0.06	0.0088780	1.00	-	-	-	-
Trp	-4.36	3.94	0.59	3.44	-1.59	184.0	0.49	0.24	-0.243375	0.70	Ala	Thr	Ala	Val
Tyr	-2.54	2.44	0.43	0.04	-1.47	147.0	0.53	0.11	-0.20751	0.92	Ala	Thr	Ala	Thr
Val	-2.59	-2.64	-1.54	-0.85	-0.02	109.0	0.61	0.13	-0.38618	0.70	-	-	-	-

The amino acid parameters used in the calculation of antigenicities and the replacement suggestions. The zz values are from Sandberg et al. (1998). The SDGly values are from Grantham (1974) and represent the structural dissimilarities of the various amino acids relative to glycine. The helix propensities are from Pace et al. (1998). The beta sheet propensities are from Street et al. (1999). The coil propensities are from Linding et al. (2003). The turn propensities are from Hutchinson et al. (1994). A dash in the replacement suggestions signifies that no change is recommended.

Table 2 - Highly pathogenic avian H5N1 transmitted to humans

```
         10         20         30         40         50         60         70
          |          |          |          |          |          |          |
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ABP51975
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ABP51968
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  AAS65615
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKKHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ABP51977
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ABO10187
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ACL11932
DQICIGSHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ABQ58979

DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ACJ68610
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ABD28180
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKLLILRDCSVAGWLLGNPMCDEF  ADG59086
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLGGVKPLILRDCSVAGWLLGNPMCDEF  ACJ68613
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  BAE46949
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ABI16504
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ACJ68612
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  AEO89030
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILGDCSVAGWLLGNPMCDEF  ACJ68607
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  AEO89021
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  AEO88980
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ACJ68608
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ACB87573
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  AEO89118
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  AEO89154
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  AEO89127
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILGDCSVAGWLLGNPMCDKF  AEO89145

DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKKHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEF  ABW90125

         80         90         100        110        120        130        140
          |          |          |          |          |          |          |
INVPEWSYIVEKANPANDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKNSWSSHEASLGVSSACPYQGK  ABP51975
INVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKSSWSSHEASLGVSSACPYQGK  ABP51968
INVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKSSWSSHEASLGVSSACPYQRK  AAS65615
INVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKSSWSSHEASLGVSSACPYQGK  ABP51977
INVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKSSWSSHEASLGVSSACLYQGQ  ABO10187
LNVPEWSYIVEKINPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGR  ACL11932
INVPEWSYIVEKASPANDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKSSWSNHEASSGVSSACPYLGR  ABQ58979

INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGT  ACJ68610
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGT  ABD28180
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGGSSACPYQGT  ADG59086
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSVCPYQGT  ACJ68613
INVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKSSWLSHEASLGVSSACPYQGK  BAE46949
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIISKSSWSDHEASSGVSSACPYQGT  ABI16504
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWPDHEASSGVSSACPYQGT  ACJ68612
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSVCPYQGT  AEO89030
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGT  ACJ68607
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIPIIPKSSWSDHEASSGVSSACPYQGT  AEO89021
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGT  AEO88980
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGT  ACJ68608
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGT  ACB87573
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGT  AEO89118
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGT  AEO89154
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGT  AEO89127
INVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGT  AEO89145

INVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKSYWSSHEASLGVSSACPYQGK  ABW90125
```

```
         150        160        170        180        190        200        210
          |          |          |          |          |          |          |
SSFFRNVVWLIKKNNAYPTIKRSYNNTNQEDLLVLWGIHHPNDAAEQTRLYQNPTTYISVGTSTLNQRLV  ABP51975
SSFFRNVVWLIKKNSTYPTIKRSYNNTNQEDLLVLWGIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLV  ABP51968
SSFFRNVVWLIKKNSTYPTIKRSYNNTNQEDLLVLWGIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLV  AAS65615
SSFFRNVVWLIKKNSTYPTIKRSYNNTNQEDLLVLWGIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLV  ABP51977
SSFFRNVVWLIKKNSTYPTIKRSYNNTNQEDLLVMWGIHHPNDAAEQIKLYQNPTTYISVGTSTLNQRLV  ABO10187
SSFFRNVVWLIKKNDAYPTIKISYNNTNQEDLLVLWGIHHPNDAAEQTRLYQNPTTYISVGTSTLNLRLV  ACL11932
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLVLWGIHHPNDEAEQIKLYQNPTTYISVGTSTLNQRLV  ABQ58979

PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNDAAEQTKLYQNPTTYISVGTSTLNQRLV  ACJ68610
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNDAAEQTKLYQNPTTYISVGTSTLNQRLV  ABD28180
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNDAAEQTKLYQNPTTYISVGTSTLNQRLV  ADG59086
PSFFRNVVWLIKKNNTYPTIKKSYNNTNQEDLLILWGIHHSNDAAEQTKLYQNPTTYISVGTSTLNQRLV  ACJ68613
SSFFRNVVWLIKKNSTYPTIKRSYNNTNQEDLLVLWGIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLV  BAE46949
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNNAAEQTKLYQNPTTYISVGTSTLNLRLV  ABI16504
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNNAAEQTKLYQNPTTYISVGTSTLNLRLV  ACJ68612
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNDAAEQTKLYQNPTTYISVGTSTLNQRLV  AEO89030
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNDAAEQTKLYQNPTTYISVGTSTLNQRLV  ACJ68607
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNNAAEQTKLYQNPTTYISVGTSTLNLRLV  AEO89021
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNDAAEQTKLYQNPTTYISVGTSTLNQRLV  AEO88980
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNDAAEQTKLYQNPTTYISVGTSTLNQRLV  ACJ68608
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQENLLILWGIHHSNDAAEQIKLYQNPTTYISVGTSTLNQRLV  ACB87573
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQENLLILWGIHHSNDAAEQIKLYQNPTTYVSIGTSTLNQRLV  AEO89118
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLVLWGIHHSNDAAEQIKLYQNPTTYVSVGTSTLNQRLV  AEO89154
PSFFRNVVWLIKKNNTYPTIKRSYNNTNQENLLILWGIHHSNDAAEQIKLYQNPTTYISVGTSTLNQRLV  AEO89127
PSFFRNVVWLTKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNNAAEQEKLYQNPTTYISVGTSTLNQRLV  AEO89145

SSFFRNVVWLTKKNSTYPTIKRSYNNTNQEDLLVLWGIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLV  ABW90125


         220        230        240        250        260        270        280
          |          |          |          |          |          |          |
PKIATRSKVNGQNGRMEFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSELEYGNCNTKCQT  ABP51975
PRIATRSKVNGQSGRMEFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSTIMKSELEYGNCNTKCQT  ABP51968
PRIATRSKVNGQSGRMEFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSTIMKSELEYGNCNTKCQT  AAS65615
PRIATRSKVNGQSGRMEFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSTIMKSELEYGNCNTKCQT  ABP51977
PRIATRSKVNGQSGRMEFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSTIMKSELEYGNCNTKCQT  ABO10187
PKIATRSKVNGQSGRMEFFWTILKPNDAINFESNGNFIAPENAYKIVKKGDSTIMKSELEYGNCNTKCQT  ACL11932
PKIATRSKVNGQSGRMEFFWAILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSELEYGNCNTKCQT  ABQ58979

PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIIKSEVEYGNCNTKCQT  ACJ68610
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIVKSEVEYGNCNTKCQT  ABD28180
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSEVEYGNCNTKCQT  ADG59086
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSEVEYGNCNTKCQT  ACJ68613
PRIATRSKVNGQSGRMEFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSTIMKSELEYGNCNTKCQT  BAE46949
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSEVEYGNCNTKCQT  ABI16504
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSEVEYGNCNTKCQT  ACJ68612
PKIATRSKVNGQSGRMDFFWTVLKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSEVEYGNCNTKCQT  AEO89030
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSEVEYGNCNTKCQT  ACJ68607
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSEVEYGNCNTKCQT  AEO89021
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSEVEYGNCNTKCQT  AEO88980
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSEVEYGNCNTKCQT  ACJ68608
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKEGDSAIMKSEVEYGNCNTKCQT  ACB87573
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKEGDSAIMKSEVEYGNCNTKCQT  AEO89118
PKIATRSKVNGQSGRMDFFWTILKSNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSEVEYGNCSTKCQT  AEO89154
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKEGDSAIMKSEVEYGNCNTKCQT  AEO89127
PKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIMKSGVEYGNCNTKCQT  AEO89145

PRIATRSKVNGQSGRMEFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSTIMKSELEYGNCNTKCQT  ABW90125
```

```
        290       300       310       320       330       340       350
         |         |         |         |         |         |         |
PMGAINSSMPFHNIHPLTIGECPKYVKSNRLVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDG  ABP51975
PMGAINSSMPFHNIHPLTIGECPKYVKSNRLVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDG  ABP51968
PMGAINSSMPFHNIHPLTIGECPKYVKSNRLVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDG  AAS65615
PMGAINSSMPFHNIHPLTIGECPKYVKSNRLVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDG  ABP51977
PMGAINSSMPFHNIHPLTIGECPKYVKSNRLVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDG  ABO10187
PVGAINSSMPFHNIHPLTIGECPKYVKSNRLVLATGLRNSPQGERRRKKRGLFGAIAGFIEGGWQGMVDG  ACL11932
PMGAINSSMPFHNIHPLTIGECPKYVKSNRLVLATGLRNAPQREGRRKKRGLFGAIAGFIEGGWQGMVDG  ABQ58979

PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  ACJ68610
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  ABD28180
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIACFIEGGWQGMVDG  ADG59086
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  ACJ68613
PMGAINSSMPFHNIHPLTIGECPKYVKSNRLVLATGLRNSPQRERRKK-RGLFGAIAGFIEGGWQGMVDG  BAE46949
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  ABI16504
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  ACJ68612
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  AEO89030
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  ACJ68607
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  AEO89021
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  AEO88980
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  ACJ68608
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  ACB87573
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  AEO89118
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  AEO89154
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRR-RGLFGAIAGFIEGGWQGMVDG  AEO89127
PIGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDG  AEO89145

PMGAINSSMPFHNIHPLTIGECPKYVKSNRLVLATGLRNSPQRET----RGLFGAIAGFIEGGWQGMVDG  ABW90125

        360       370       380       390       400       410       420
         |         |         |         |         |         |         |
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ABP51975
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ABP51968
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  AAS65615
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ABP51977
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ABO10187
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ACL11932
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAIGREFNNLERRIENLNKKMEDGFLD  ABQ58979

WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ACJ68610
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ABD28180
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ADG59086
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ACJ68613
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  BAE46949
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ABI16504
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ACJ68612
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  AEO89030
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ACJ68607
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  AEO89021
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  AEO88980
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ACJ68608
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ACB87573
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  AEO89118
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  AEO89154
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  AEO89127
WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  AEO89145

WYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD  ABW90125
```

```
         430        440        450        460        470        480
          |          |          |          |          |          |
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ABP51975
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ABP51968
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      AAS65615
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ABP51977
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ABO10187
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHRCDNECME      ACL11932
VWTYNAELLVLMENERTLDFHDSNVKNLYEKVRLQLRDNAKGLGNGCFEFYHKCDNECME      ABQ58979

VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ACJ68610
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ABD28180
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ADG59086
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ACJ68613
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      BAE46949
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ABI16504
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ACJ68612
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      AEO89030
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ACJ68607
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      AEO89021
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      AEO88980
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ACJ68608
VWTYNAELLVLMENERTLDFHDSNVKNLYDRVRLQLRDNAKELGNGCFEFYHKCDNECME      ACB87573
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      AEO89118
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      AEO89154
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      AEO89127
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      AEO89145

VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ABW90125

         490        500
          |          |
SVRNGTYDYPQYSEEARLKREEI                                            ABP51975
SVRNGTYDYPQYSEEARLKREEI                                            ABP51968
SVRNGTYDYPQYSEEARLKREEI                                            AAS65615
SVRNGTYDYPQYSEEARLKREEI                                            ABP51977
SVRNGTYDYPQYSEEARLKREEI                                            ABO10187
SVRNGTYDYPQYSEESRLKREEI                                            ACL11932
SVKNGTYDYPQYSEEATLNREEI                                            ABQ58979

SVRNGTYDYPQYSEEARLKREEI                                            ACJ68610
SVRNGTYDYPQYSEEARLKREEI                                            ABD28180
SVRNGTYDYPQYSEEARLKREEI                                            ADG59086
SVRNGTYDYPQYSEEARLKREEI                                            ACJ68613
SVRNGTYDYPQYSEEAKLKRGEI                                            BAE46949
SVRNGTYDYPQYSEEARLKREEI                                            ABI16504
SVRNGTYDYPQYSEEARLKREEI                                            ACJ68612
SVRNGTYDYPQYSEEARLKREEI                                            AEO89030
SVRNGTYDYPQYSEEARLKREEI                                            ACJ68607
SVRNGTYDYPQYSEEARLKREEI                                            AEO89021
SVRNGTYDYPQYSEEARLKREEI                                            AEO88980
SVRNGTYDYPQYSEEARLKREEI                                            ACJ68608
SVRNGTYDYPQYSEEARLKREEI                                            ACB87573
SVRNGTYDYPQYSEEARLKREEI                                            AEO89118
SVRNGTYDYPQYSEEARLKREEI                                            AEO89154
SVRNGTYDYPQYSEEARLKREEI                                            AEO89127
SVRNGTYDYPQYSEEARLKREEI                                            AEO89145

SVRNGTYDYPQYSEEARLKREEI                                            ABW90125
```

List of the GenBank Entries included in this study with source location and year of deposition, and number of identical sequences found (in parentheses). The entries chosen to represent those with different cleavage-site sizes are shown bold.

ABP51975 A/Hong Kong/213/2003 2003//	**(6)**
ABP51968 A/Prachinburi/6231/2004 2004//	(1)
AAS65615 A/Thailand/1(KAN-1)/2004 2004//	(10)
ABP51977 A/Viet Nam/1203/2004 2004//	(23)
ABO10187 A/Cambodia/JP52a/2005 2005//	(1)
ACL11932 A/Bangladesh/207095/2008 2008/02/	(1)
ABQ58979 A/Beijing/01/2003 2003//	(1)
ACJ68610 A/Fujian/1/2005 2005/12/06	**(2)**
ABD28180 A/Anhui/1/2005 2005/11/01	(6)
ADG59086 A/Guangxi/1/2005 2005//	(3)
ACJ68613 A/Jiangxi/1/2005 2005/12/04	(1)
BAE46949 A/noi/30408/2005 2005//	(1)
ABI16504 A/China/GD01/2006 2006//	(1)
ACJ68612 A/Guangdong/01/2006 2006/03/04	(3)
AEO89030 A/Hubei/1/2006 2006/04/01	(2)
ACJ68607 A/Hunan/1/2006 2006/01/27	(3)
AEO89021 A/Sngi/1/2006 2006/03/13	(2)
AEO88980 A/Sichuan/2/2006 2006/01/10	(2)
ACJ68608 A/Zhejiang/1/2006 2006/02/21	(1)
ACB87573 A/Jiangsu/1/2007 2007//	(2)
AEO89118 A/Beijing/1/2009 2008/12/04	(2)
AEO89154 A/Guizhou/1/2009 2009/01/15	(1)
AEO89127 A/Sndong/1/2009 2009/01/05	(1)
AEO89145 A/Xinjiang/1/2009 2009/01/10	(2)
ABW90125 A/Viet Nam/1203/2004 2004//	**(13)**

Table 3: Cleavage-site regions of hemagglutinins of highly pathogenic H5N1 avian viruses transmitted to humans

```
         320        330        340        350        360        370        380
          |          |          |          |          |          |          |
LVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ABP51975
LVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ABP51968
LVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN AAS65615
LVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ABP51977
LVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ABO10187
LVLATGLRNSPQGERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ACL11932
LVLATGLRNAPQREGRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ABQ58979

LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ACJ68610
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ABD28180
LVLATGLRNSPLRERRRK-RGLFGAIACFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ADG59086
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ACJ68613
LVLATGLRNSPQRERRKK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN BAE46949
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ABI16504
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ACJ68612
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN AEO89030
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ACJ68607
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN AEO89021
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN AEO88980
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ACJ68608
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ACB87573
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN AEO89118
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN AEO89154
LVLATGLRNSPLRERRRR-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN AEO89127
LVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN AEO89145

LVLATGLRNSPQRET----RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTN ABW90125
```

The cleavage-site regions of the HPAI H5N1 sequences that we analyzed are listed showing the differences in the number of residues in the cleavage site. Sequence numbering follows that in Table 2.

the method of Connolly (1983) using programs developed by Sheriff et al. (1985).

There is a variety in size and shape of epitopes so that the exact extent of an epitope can only be guessed. A survey made by one us in 2012 (Padlan, unpublished) on antibody-antigen complexes available at the time in PDB showed that almost all of the residues in the antigenic sites are within 17 Angstroms of the antigenic-site center (see, Figure 1). Here, we assume that the putative cleavage-site epitope of HPAI H5N1 is also comprised of all the residues that are within 17 Angstroms of the alpha carbon of the cleavage-site arginine at position 330 of the hemagglutinin.

Trial calculations on the degree of de-Antigenization using various exposure levels were performed in an attempt to find the one that is probably best for the design of the desired subunit vaccines. The effect of lowering the reactivity of all the residues in the molecule, except for those in the putative cleavage-site epitope, was calculated with the side-chains whose fractional exposures to solvent are 0.50 (half exposed), 0.3333 (one-third exposed), 0.25 (one-fourth exposed), and 0.20 (one-fifth exposed). The exposure levels are used as weights in the contribution of a particular residue to the antigenicity of an epitope. Mutations to lower the reactivity at the individual positions followed the replacement rules shown in Table 1. The antigenicity computed for an epitope is then expressed in terms of standard deviations above the mean for all the putative epitopes of the structure.

RESULTS

The results of our de-Antigenization analyses are summarized in Tables 4, 5, and 6. The de-Antigenized sequences at the different levels of exposure are labeled ""mut "exposure level"" in the tables. Our preference for possible subunit vaccines against the three representative HPAI H5N1 virus types are those resulting from de-Antigenization at 0.20 exposure level; those are shown in red and labelled "mut 0.20" in the tables. Other levels of de-Antigenization could also be used. Since it is the hemagglutinin that initially binds to the target cell, all the resulting changes in this fragment should be incorporated in the proposed vaccines. The "mut" sequences in Tables 4, 5, and 6 could be used in the development of potentially useful subunit vaccines, or as the basis for transformed viral vaccines.

Examination of the cleavage-site regions reveals the existence of cleavage-site regions that are common to viruses found in several countries. Of the 1271 complete HPAI H5N1 sequences that we obtained from the NCBI Influenza Virus Database, the most common is the cleavage-site region of ABP51975 from Hong Kong which is shared by AAT73266 from Thailand, by AAX53505 from China, by ABU99029 from Indonesia, and by AAT73277 from Viet Nam, i.e., by 405 of the sequences. Another cleavage-site region that is common is that of ACU24777 from Bangladesh, which is shared by ABE68921 from China and by ABY76247 from Egypt, i.e., by 225 of the sequences. The ABP51975 vaccine alone might be able to protect against one third of possible cases, while the simultaneous use of the ABP51975 and ABY76247 vaccines might be able to protect against half. Further, most of the more reactive residues are common among the various cleavage-site regions and several of the differences are conservative substitutions, e.g., lysine vs. arginine, so that the ABP51975/ABY76247 vaccine combination may already be protective against most, if not all, HPAI H5N1 strains.

A possible vaccine against the "ABY76247 family" of HPAI H5N1 viruses could be proposed using the vaccine we have designed against the "ABP51975-family" of viruses. The ABY76247 hemagglutinin sequence differs from that of ABP51975 at 15 positions: ABY76247 has a Leu for the Ile at position 71 (numbering follows that in Table 2), Ile for the Ala at position 83, Asn for the Asp at position 94, Ser for the Asn at position 120, Asp for the Ser at position 124, Ser for the Leu at position 129, Arg for the Lys at position 140, Asp for the Asn at position 154, Ser for the Asn at position 223, Ser for the Pro at position 235, Asn for the Tyr at position 252, Thr for the Ala at position 263, Ile for the Met at position 282, Gly for the Arg at position 323 in the cleavage-site region, and Arg for the Lys at position 473. These 15 differences in a sequence 503 residues should not significantly affect the overall structure of the molecule and a ABY76247 vaccine tailored after the ABP51975 vaccine can be proposed. Following the replacements rules shown in Table 1, a possible ABY76247 vaccine is shown in Table 7.

CONCLUSION

The advantage of our de-Antigenized subunit vaccines over those which use unaltered subunits is the fact that the latter can be expected to produce antibodies against all the epitopes of the protein, whereas our proposed subunit vaccines would predominantly promote the production of antibodies against the cleavage sites which could be expected to prevent the structural change in hemagglutinin that is needed for entry into target cells.

In our earlier study on seasonal H3N2, pandemic H1N1, and seasonal B influenza (Mikita and Padlan 2021), we found that de-Antigenization altered all the antigenic and receptor-binding sites that had been identified by others in H3, H1 and B hemagglutinin (Caton et al. 1982, Gamblin et al. 2004, Ha et al. 2001, Sauter et al. 1992, Smith et al. 2004, Stray and Pittman 2012). That finding led us to conclude that the de-Antigenization of the seasonal H3, pandemic H1, and seasonal B hemagglutinins would most probably direct the antibody response mainly to the chosen target, i.e., the cleavage site, and nowhere else. We expect the same for HPAI H5N1.

Although we have concentrated on the H5N1 HPAI, vaccines against other pathogenic avian influenza variants could be similarly designed by this method.

Cleavage of the hemagglutinin is required for the infectivity of the virus so that preventing cleavage is a possible means of preventing viral infection. We had proposed that antibodies purposely directed at the cleavage site could be used as subunit vaccines against seasonal influenza (Mikita and Padlan 2011). We extend this suggestion to highly pathogenic avian influenza. In their study of the structure of hemagglutinin, Wilson et al. (1981) replaced the cleavage-site arginine with glutamine, which prevented cleavage and which resulted in an intact molecule suitable for their crystallographic study. The judicious replacement of the cleavage-site arginine to prevent cleavage is thus another way of producing subunit vaccines against influenza. This is being pursued for influenza B (Stech et. al. 2011, Saelens 2012).

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest. The views expressed in this article are those of the authors and do not reflect the official policy of the Department of Army/Navy/Air Force, Department of Defense, or U.S. Government.

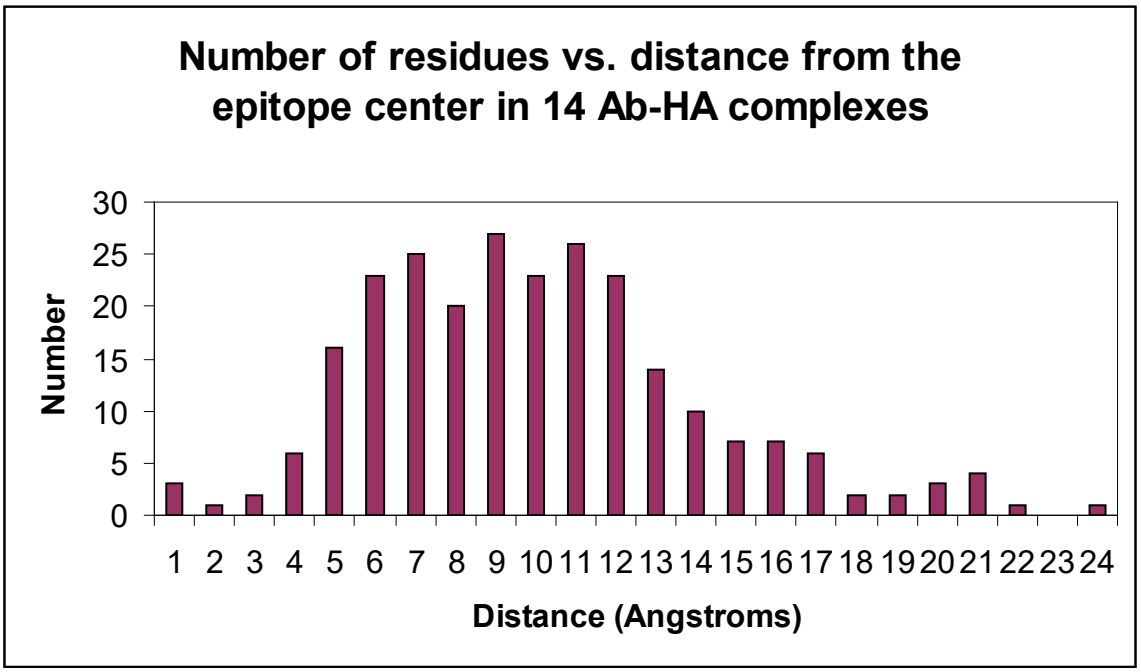

Figure 1: Survey on antibody-antigen complexes

Plot of the number of residues and their distances from the cleavage site of hemagglutinin computed from 14 antibody-hemagglutinin complexes in the Protein Data Bank (entries: 3ZTN, 3ZTJ, 3SM5, 3SDY, 3LZF, 3GBN, 3GBM, 3FKU, 1KEN, 1EO8, 1QFU, 2VIT, 2VIS, 2VIR) that were available at the time of the study (Padlan, unpublished).

Table 4: ABP51975 (a highly pathogenic avian H5N1 transmitted to humans)

```
         10         20         30         40         50         60
          |          |          |          |          |          |
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGW      ABP51975
DQICIGYHANNSTEQVSTIMETGVTVTTATDILEKTHNGTLCDLGGVTPLILAGCSVAGW      mut 0.50
DTICIGYHANNSTATVSTIMATGVTVTTATTILETTTSGTLCDLGGVTPLILAGCSVAGW      mut 0.3333
STICIGYHANNSTATVSTIMATGVTVTTATTILETTTSGTLCDLGGVTPLILAGCSVAGW      mut 0.25
STICIGYHANNSTATVSTIMATGVTVTTATTILTTTTSGTLCTLGGVTPLILAGCSVAGA      mut 0.20
         70         80         90         100        110        120
          |          |          |          |          |          |
LLGNPMCDEFINVPEWSYIVEKANPANDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKN      ABP51975
LLGNPMCSAAIGVPEWSYIVEKAGPANGLCYPGTFGDYEELKHLLSRINHFEKIQIIPAA      mut 0.50
LLGNPMCSAAIGVPTWSYIVEKAGPATGLCYPGTFGDYEELKHLLSAINHFTTITIIPAA      mut 0.3333
LLGNPMCSAAIGVPTWSYIVETAGPATGLCYPGTFGDAAELKHLLSAITTFTTITIIPAA      mut 0.25
LLGNPMCSAAIGVPTWSYIVETAGPATGLCYPGTFGDAAELKHLLSAITTFTTITIIPAA      mut 0.20
         130        140        150        160        170        180
          |          |          |          |          |          |
SWSSHEASLGVSSACPYQGKSSFFRNVVWLIKKNNAYPTIKRSYNNTNQEDLLVLWGIHH      ABP51975
SWSSHEASLGVSSACPTTGTSSFFRNVVWLIKTGGAYPTITRSYTNTNTEDLLVLWGIHH      mut 0.50
SWSSHTASLGVSSACPTTGTSSFFANVVWLIKTGGAYPTITTSYTNTNTTSLLVLWGIHH      mut 0.3333
SWSSHTASLGVSSACPTTGTSSFFANVVWLIKTGGAYPTITTSTTNTGTTSLLVLWGIHT      mut 0.25
SWSSHTASLGVSSACPTTGTSSFFANVVWLITTGGAYPTITTSTTNTGTTSLLVLWGIHT      mut 0.20
         190        200        210        220        230        240
          |          |          |          |          |          |
PNDAAEQTRLYQNPTTYISVGTSTLNQRLVPKIATRSKVNGQNGRMEFFWTILKPNDAIN      ABP51975
PGGAAEQTALYQNPTTYISVGTSTLNQRLVPTIATRSTVGGQNGRMEFFWTILTPGDAIN      mut 0.50
PGGAAAQTALYASPTTYISVGTSTLNQRLVPTIATRSTVGGQGGRMEFFWTILTPGSAIT      mut 0.3333
PGGAAAQTALYASPTTAISVGTSTLTTRLVPTIATRSTVGGTGGRMTFTWTILTPGSAIT      mut 0.25
PGGAAAQTALYASPTTAISVGTSTLTTRLVPTIATRSTVGGTGGRMTFTWTILTPGSAIT      mut 0.20
         250        260        270        280        290        300
          |          |          |          |          |          |
FESNGNFIAPEYAYKIVKKGDSAIMKSELEYGNCNTKCQTPMGAINSSMPFHNIHPLTIG      ABP51975
FESNGNFIAPEYAYKIVTTGGSAIMKSALAYGTCTTTCQTPMGAISSSMPFHNIHPLTIG      mut 0.50
FESGGNFIAPEYAYKIVTTGGSAIMTSALATGTCTTTCQTPMGAISSSMPFHSITPLTIG      mut 0.3333
```

```
FESGGNFIAPETAYKIVTTGGSAIMTSALATGTCTTTCQTPMGAISSSMPFHSITPLTIG      mut 0.25
FTSGGNFIAPTTAYKIVTTGGSAIMTSALATGTCTTTCQTPMGAISSSMPFHSITPLTIG      mut 0.20
        310       320       330       340       350       360
         |         |         |         |         |         |
ECPKYVKSNRLVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQ      ABP51975
TCPKYVTSGALVLATGLRNSPQRERRRKRGLFGAIAGFIEGGWQGMVGGWYGYHTSNEQ      mut 0.50
TCPKYVTSGALVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVGGWYGYHTSNTT      mut 0.3333
TCPTYVTSGALVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVGGVYGYHTSNTT      mut 0.25
TCPTYVTSGALVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVGGVYGYHTSNTT      mut 0.20
        370       380       390       400       410       420
         |         |         |         |         |         |
GSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD      ABP51975
GSGYAADAASTAAAIAGVTNKVASIIAKMGTQFEAVGRAFSGLERRIENLNAKMEAGFLA      mut 0.50
GSGYAADAASTAAAIAGVTAKVASIIAKMGTQFAAVGRAFSGLERRIANLNAAMEAGFLA      mut 0.3333
GSGTAADAASTAAAIAGVTAAVASIIAKMGTQFAAVGRAFSGLAARIANLNAAMAAGFLA      mut 0.25
GSGTAADAASTAAAIAGVTAAVASIIAKMGTQFAAVGRAFSGLAAAIANLAAAMAAGFLA      mut 0.20
        430       440       450       460       470       480
         |         |         |         |         |         |
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ABP51975
VWTYNAALLVLMENEATLAAHASNVAALYAKVRLQLRANAKELGNGCFTFYHTCAAACMA      mut 0.50
VWTYAAALLVLMENAATLAAHASAVAALYAAVRLQLRANATELGNGCFTFTHTCAAACMA      mut 0.3333
VWTYAAALLVLMENAATLAAHASAVAALYAAVRLQLRANATTLGNGCFTFTHTCAAACMA      mut 0.25
VWTYAAALLVLMANAATLAAHASAVAALAAAVRLQLAANATTLGNGCFTFTHTCAAACMA      mut 0.20
        490       500
         |         |
SVRNGTYDYPQYSEEARLKREEI                                           ABP51975
SVRGGTYSYPAYSAAAALARAAI                                           mut 0.50
SVAGGTYSYPAYSAAAALAAAAI                                           mut 0.3333
SVAGGTYSAPATSAAAALAAAAI                                           mut 0.25
SVAGGTYSAPATSAAAALAAAAI                                           mut 0.20
```

The result of de-Antigenization of GenBank Entry ABP51975 at various levels of exposure is shown. The residues in ABP51975 that are within 17 Angstroms of the cleavage-site arginine are underlined. The residues whose contribution to the antigenicity of the putative cleavage-site epitope is at least 2 standard deviations above the mean for the whole structure are shown bold.

Table 5: ACJ68610 (a highly pathogenic avian H5N1 transmitted to humans)

```
        10        20        30        40        50        60
         |         |         |         |         |         |
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGW      ACJ68610
DQICIGYHANNSTEQVSTIMETGVTVTTATDILEKTHNGTLCDLTGVTPLILATCSVAGW      mut 0.50
DAICIGYHANNSTAAVSTIMATGVTVTTATGILETTTTGTLCDLTGVTPLILATCSVAGW      mut 0.3333
SAICIGYHANNSTAAVSTIMATGVTVTTATGILETTTTGTLCDLTGVTPLILATCSVAGW      mut 0.25
SAICIGYHANNSTAAVSTIMATGVTVTTATGILTTTTTGTLCGLTGVTPLILATCSVAGA      mut 0.20
        70        80        90        100       110       120
         |         |         |         |         |         |
LLGNPMCDEFINVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKS      ACJ68610
LLGNPMCSTAISVPEWSYIVEKATPANGLCYPGGFSDYEELKHLLSRINHFETIQIIPTS      mut 0.50
LLGNPMCSTAISVPTWSYIVEKATPAGGLCYPGGFSDYEELKHLLSAINHFTTITIIPTS      mut 0.3333
LLGNPMCSTAISVPTWSYIVETATPAGGLCYPGGFSDTTELKHLLSAIATFTTITIIPTS      mut 0.25
LLGNPMCSTAISVPTWSYIVETATPAGGLCYPGGFSDTTELKHLLSAIATFTTITIIPTS      mut 0.20
        130       140       150       160       170       180
         |         |         |         |         |         |
SWSDHEASSGVSSACPYQGTPSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHH      ACJ68610
SWSAHEASSGVSSACPTTGTPSFFRNVVWLIKKSTTYPTITRSYTNTNTEDLLILWGIHH      mut 0.50
SWSAHAASSGVSSACPTTGTPSFFTNVVWLIKTSTTYPTITASYTNTNTASLLILWGIHH      mut 0.3333
SWSAHAASSGVSSACPTTGTPSFFTNVVWLIKTSTTYPTITASTTNTTTASLLILWGIHT      mut 0.25
SWSAHAASSGVSSACPTTGTPSFFTNVVWLITTSTTYPTITASTTNTTTASLLILWGIHT      mut 0.20
        190       200       210       220       230       240
         |         |         |         |         |         |
SNDAAEQTKLYQNPTTYISVGTSTLNQRLVPKIATRSKVNGQSGRMDFFWTILKPNDAIN      ACJ68610
STTAAEQTALYQNPTTYISVGTSTLNQRLVPTIATRSTVSGQSGRMDFFWTILTPTDAIN      mut 0.50
STTAATQTALYAAPTTYISVGTSTLNQRLVPTIATRSTVSGQSGRMDFFWTILTPTTAIG      mut 0.3333
STTAATQTALYAAPTTYISVGTSTLNTRLVPTIATRSTVSGASGRMDFTWTILTPTTAIG      mut 0.25
```

```
STTAATQTALYAAPTTAISVGTSTLSTRLVPTIATRSTVSGASGRMGFTWTILTPTTAIG      mut 0.20
        250       260       270       280       290       300
         |         |         |         |         |         |
FESNGNFIAPEYAYKIVKKGDSAIIKSEVEYGNCNTKCQTPIGAINSSMPFHNIHPLTIG      ACJ68610
FESNGNFIAPEYAYKIVTTGTSAIIKSTVTYGSCTTTCQTPIGAIGSSMPFHNIHPLTIG      mut 0.50
FESTGNFIAPEYAYKIVTTGTSAIITSTVTAGSCTTTCQTPIGAIGSSMPFHGITPLTIG      mut 0.3333
FESTGNFIAPETAYKIVTTGTSAIITSTVTAGSCTTTCQTPIGAIGSSMPFHGITPLTIG      mut 0.25
FTSTGNFIAPTTAYKIVTTGTSAIITSTVTAGSCTTTCQTPIGAIGSSMPFHGITPLTIG      mut 0.20
        310       320       330       340       350       360
         |         |         |         |         |         |
ECPKYVKSNKLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQ      ACJ68610
ACPKYVTSTTLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVGGWYGYHTSNEQ      mut 0.50
ACPKYVTSTTLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVGGWYGYHTSNTT      mut 0.3333
ACPTYVTSTTLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVGGAYGYHTSNTT      mut 0.25
ACPTYVTSTTLVLATGLRNSPLRERRRK-RGLFGAIAGFIEGGWQGMVGGAYGYHTSNTT      mut 0.20
        370       380       390       400       410       420
         |         |         |         |         |         |
GSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD      ACJ68610
GSGYAADTTSTAAAIAGVTNKVASIIAKMATQFEAVGRAFSSLERRIENLNAKMEAGFLA      mut 0.50
GSGTAADTTSTAAAIAGVTAKVASIIAKMATQFTAVGRAFSSLERRIANLNAAMEAGFLA      mut 0.3333
GSGTAADTTSTAAAIAGVTAAVASIIAKMATQFTAVGRAFSSLAARIANLNAAMAAGFLA      mut 0.25
GSGTAADTTSTAAAIAGVTAAVASIIAKMATQFTAVGRAFSSLAARIANLAAAMAAGFLA      mut 0.20
        430       440       450       460       470       480
         |         |         |         |         |         |
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ACJ68610
VWTYNAALLVLMENEATLAAHASNVAALYAKVRLQLRANAKELGNGCFTFYHTCGGACMA      mut 0.50
VWTYAAALLVLMENAATLAAHASAVAALYAAVRLQLRANAAELGNGCFTFYHTCGGACMA      mut 0.3333
VWTYAAALLVLMENAATLAAHASAVAALYAAVRLQLRANAAALGNGCFTFTHTCGGACMA      mut 0.25
VWTYAAALLVLMANAATLAAHASAVAALAAAVRLQLAANAAALGNGCFTFTHTCGGACMA      mut 0.20
        490       500
         |         |
SVRNGTYDYPQYSEEARLKREEI                                           ACJ68610
SVRAGTYGYPAYSEAAALARAAI                                           mut 0.50
SVAAGTYGYPAYSAAAALAAAAI                                           mut 0.3333
SVAAGTYGTPAASAAAALAAAAI                                           mut 0.25
SVAAGTYGTPAASAAAALAAAAI                                           mut 0.20
```

The result of de-Antigenization of GenBank Entry ACJ68610 at various levels of exposure is shown. The residues in ACJ68610 that are within 17 Angstroms of the cleavage-site arginine are underlined. The residues whose contribution to the antigenicity of the putative cleavage-site epitope is at least 2 standard deviations above the mean for the whole structure are shown bold.

Table 6: ABW90125 (a highly pathogenic avian H5N1 transmitted to humans)

```
        10        20        30        40        50        60
         |         |         |         |         |         |
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKKHNGKLCDLDGVKPLILRDCSVAGW      ABW90125
DQICIGYHANNSTAQVSTIMETGVTVTTATDILEKTHNGTLCDLTGVTPLILATCSVAGW      mut 0.50
DAICIGYHANNSTAAVSTIMATGVTVTTATGILETTTTGTLCDLTGVTPLILATCSVAGW      mut 0.3333
SAICIGYHANNSTAAVSTIMATGVTVTTATGILTTTTTGTLCDLTGVTPLILATCSVAGW      mut 0.25
SAICIGYHANNSTAAVSTIMATGVTVTTATGILTTTTTGTLCGLTGVTPLILATCSVAGA      mut 0.20
        70        80        90       100       110       120
         |         |         |         |         |         |
LLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKS      ABW90125
LLGNPMCSEAISVPEWSYIVEKATPVNGLCYPGGFSDYEELKHLLSRINHFETIQIIPTS      mut 0.50
LLGNPMCSTAISVPTWSYIVEKATPVGGLCYPGGFSDYEELKHLLSAINHFTTITIIPTS      mut 0.3333
LLGNPMCSTAISVPTWSYIVETATPVGGLCYPGGFSDTTELKHLLSAIATFTTITIIPTS      mut 0.25
LLGNPMCSTAISVPTWSYIVETATPVGGLCYPGGFSDTTELKHLLSAIATFTTITIIPTS      mut 0.20
        130       140       150       160       170       180
         |         |         |         |         |         |
YWSSHEASLGVSSACPYQGKSSFFRNVVWLTKKNSTYPTIKRSYNNTNQEDLLVLWGIHH      ABW90125
TWSSHEASLGVSSACPTTGTSSFFRNVVWLTKKSSTYPTITRSYTNTNTEDLLVLWGIHH      mut 0.50
TWSSHAASLGVSSACPTTGTSSFFTNVVWLTKTSSTYPTITASYTNTNTASLLVLWGIHH      mut 0.3333
TWSSHAASLGVSSACPTTGTSSFFTNVVWLTTTSSTYPTITASTTNTTTASLLVLWGIHH      mut 0.25
TWSSHAASLGVSSACPTTGTSSFFTNVVVLTTTSSTYPTITASTTNTTTASLLVLWGIHH      mut 0.20
        190       200       210       220       230       240
```

```
         |         |         |         |         |         |
PNDAAEQTKLYQNPTTYISVGTSTLNQRLVPRIATRSKVNGQSGRMEFFWTILKPNDAIN       ABW90125
PTTAAEQTALYQAPTTYISVGTSTLNQRLVPRIATRSTVSGQSGRMEFFWTILTPTDAIN       mut 0.50
PTTAATQTALYAAPTTYISVGTSTLNQRLVPTIATRSTVSGQSGRMEFFWTILTPTTAIG       mut 0.3333
PTTAATQTALYAAPTTYISVGTSTLSTRLVPTIATRSTVSGASGRMTFTWTILTPTTAIG       mut 0.25
PTTAATQTALYAAPTTAISVGTSTLSTRLVPTIATRSTVSGASGRMTFTWTILTPTTAIG       mut 0.20
        250       260       270       280       290       300
         |         |         |         |         |         |
FESNGNFIAPEYAYKIVKKGDSTIMKSELEYGNCNTKCQTPMGAINSSMPFHNIHPLTIG       ABW90125
FESNGNFIAPEYAYKIVTTGTSTIMKSTLTYGSCTTTCQTPMGAIGSSMPFHNIHPLTIG       mut 0.50
FESTGNFIAPEYAYKIVTTGTSTIMTSTLTAGSCTTTCQTPMGAIGSSMPFHGITPLTIG       mut 0.3333
FESTGNFIAPETAYKIVTTGTSTIMTSTLTAGSCTTTCQTPMGAIGSSMPFHGITPLTIG       mut 0.25
FTSTGNFIAPTTAYKIVTTGTSTIMTSTLTAGSCTTTCQTPMGAIGSSMPFHGITPLTIG       mut 0.20
        310       320       330       340       350       360
         |         |         |         |         |         |
ECPKYVKSNRLVLATGLRNSPQRET----RGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQ       ABW90125
ACPKYVTSTALVLATGLRNSPQRET----RGLFGAIAGFIEGGWQGMVGGWYGYHTSNEQ       mut 0.50
ACPKYVTSTALVLATGLRNSPQRET----RGLFGAIAGFIEGGWQGMVGGWYGYHTSNTT       mut 0.3333
ACPTYVTSTALVLATGLRNSPQRET----RGLFGAIAGFIEGGWQGMVGGAYGYHTSNTT       mut 0.25
ACPTYVTSTALVLATGLRNSPQRET----RGLFGAIAGFIEGGWQGMVGGAYGYHTSNTT       mut 0.20
        370       380       390       400       410       420
         |         |         |         |         |         |
GSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD       ABW90125
GSGYAADTTSTAAAIAGVTNKVASIIAKMATQFEAVGRAFSSLERRIENLNAKMEAGFLA       mut 0.50
GSGTAADTTSTAAAIAGVTAKVASIIAKMATQFTAVGRAFSSLERRIANLNAAMEAGFLA       mut 0.3333
GSGTAADTTSTAAAIAGVTAAVASIIAKMATQFTAVGRAFSSLAARIANLNAAMAAGFLA       mut 0.25
GSGTAADTTSTAAAIAGVTAAVASIIAKMATQFTAVGRAFSSLAARIANLAAAMAAGFLA       mut 0.20
        430       440       450       460       470       480
         |         |         |         |         |         |
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME       ABW90125
VWTYNAALLVLMENEATLAAHASNVAALYAKVRLQLRANAKELGNGCFTFYHTCGGACMA       mut 0.50
VWTYAAALLVLMENAATLAAHASAVAALYAAVRLQLRANAAELGNGCFTFTHTCGGACMA       mut 0.3333
VWTYAAALLVLMENAATLAAHASAVAALYAAVRLQLRANAAALGNGCFTFTHTCGGACMA       mut 0.25
VWTYAAALLVLMANAATLAAHASAVAALAAAVRLQLAANAAALGNGCFTFTHTCGGACMA       mut 0.20
        490       500
         |         |
SVRNGTYDYPQYSEEARLKREEI                                            ABW90125
SVRAGTYGYPAYSEAARLARAAI                                            mut 0.50
SVAAGTYGYPAYSAAAALAAAAI                                            mut 0.3333
SVAAGTYGTPAASAAAALAAAAI                                            mut 0.25
SVAAGTYGTPAASAAAALAAAAI                                            mut 0.20
```

The result of de-Antigenization of GenBank Entry ABW90125 at various levels of exposure is shown. The residues in ABW90125 that are within 17 Angstroms of the cleavage-site arginine are underlined. The residues whose contribution to the antigenicity of the putative cleavage-site epitope is at least 2 standard deviations above the mean for the whole structure are shown bold.

Table 7: ABP51975 and ABY76247 possible vaccines

```
         10        20        30        40        50        60
          |         |         |         |         |         |
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGW       ABP51975
DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGW       ABY76247
STICIGYHANNSTATVSTIMATGVTVTTATTILTTTTSGTLCTLGGVTPLILAGCSVAGA       mut 0.20
STICIGYHANNSTATVSTIMATGVTVTTATTILTTTTSGTLCTLGGVTPLILAGCSVAGA       mut*
         70        80        90        100       110       120
          |*        |  *      |   *     |         |         *
LLGNPMCDEFINVPEWSYIVEKANPANDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKN       ABP51975
LLGNPMCDEFLNVPEWSYIVEKINPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKS       ABY76247
LLGNPMCSAAIGVPTWSYIVETAGPATGLCYPGTFGDAAELKHLLSAITTFTTITIIPAA       mut 0.20
LLGNPMCSAALGVPTWSYIVETIGPATGLCYPGNFGDAAELKHLLSAITTFTTITIIPAS       mut*
         130       140       150       160       170       180
    *    *|         *         |   *     |         |         |
SWSSHEASLGVSSACPYQGKSSFFRNVVWLIKKNNAYPTIKRSYNNTNQEDLLVLWGIHH       ABP51975
SWSDHEASSGVSSACPYQGRSSFFRNVVWLIKKDNAYPTIKRSYNNTNQEDLLVLWGIHH       ABY76247
```

```
SWSSHTASLGVSSACPTTGTSSFFANVVWLITTGGAYPTITTSTTNTGTTSLLVLWGIHT      mut 0.20
SWSGHTASSGVSSACPTTGASSFFANVVWLITTGGAYPTITTSTTNTGTTSLLVLWGIHT      mut*
        190       200       210       220       230       240
         |         |         |         |  *      |    *    |
PNDAAEQTRLYQNPTTYISVGTSTLNQRLVPKIATRSKVNGQNGRMEFFWTILKPNDAIN      ABP51975
PNDAAEQTRLYQNPTTYISVGTSTLNQRLVPKIATRSKVNGQSGRMEFFWTILKSNDAIN      ABY76247
PGGAAAQTALYASPTTAISVGTSTLTTRLVPTIATRSTVGGTGGRMTFTWTILTPGSAIT      mut 0.20
PGGAAAQTALYASPTTAISVGTSTLTTRLVPTIATRSTVGGTSGRMTFTWTILTSGSAIT      mut*
        250       260       270       280       290       300
         | *       |  *      |         | *       |         |
FESNGNFIAPEYAYKIVKKGDSAIMKSELEYGNCNTKCQTPMGAINSSMPFHNIHPLTIG      ABP51975
FESNGNFIAPENAYKIVKKGDSTIMKSELEYGNCNTKCQTPIGAINSSMPFHNIHPLTIG      ABY76247
FTSGGNFIAPTTAYKIVTTGGSAIMTSALATGTCTTTCQTPMGAISSSMPFHSITPLTIG      mut 0.20
FTSGGNFIAPTNAYKIVTTGGSTIMTSALATGTCTTTCQTPIGAISSSMPFHSITPLTIG      mut*
        310       320       330       340       350       360
         |         |  *      |         |         |         |
ECPKYVKSNRLVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQ      ABP51975
ECPKYVKSNRLVLATGLRNSPQGERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQ      ABY76247
TCPTYVTSGALVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVGGVYGYHTSNTT      mut 0.20
TCPTYVTSGALVLATGLRNSPQGERRRKKRGLFGAIAGFIEGGWQGMVGGVYGYHTSNTT      mut*
        370       380       390       400       410       420
         |         |         |         |         |         |
GSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD      ABP51975
GSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLD      ABY76247
GSGTAADAASTAAAIAGVTAAVASIIAKMGTQFAAVGRAFSGLAAAIANLAAAMAAGFLA      mut 0.20
GSGTAADAASTAAAIAGVTAAVASIIAKMGTQFAAVGRAFSGLAAAIANLAAAMAAGFLA      mut*
        430       440       450       460       470       480
         |         |         |         |         |  *      |
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECME      ABP51975
VWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHRCDNECME      ABY76247
VWTYAAALLVLMANAATLAAHASAVAALAAAVRLQLAANATTLGNGCFTFTHTCAAACMA      mut 0.20
VWTYAAALLVLMANAATLAAHASAVAALAAAVRLQLAANATTLGNGCFTFTHACAAACMA      mut*
        490       500
         |         |
SVRNGTYDYPQYSEEARLKREEI                                           ABP51975
SVRNGTYDYPQYSEEARLKREEI                                           ABY76247
SVAGGTYSAPATSAAAALAAAAI                                           mut 0.20
SVAGGTYSAPATSAAAALAAAAI                                           mut*
```

The hemagglutinin sequences of ABP51975 and ABY76247 (dark blue) are shown. The "mut" sequence (shown in red) is for ABP51975 after de-Antigenization at exposure level 0.20 (see Table 4). The positions where ABY76247 differs from ABP51975 are indicated by asterisks (*) above the sequences. The proposed ABY76247 vaccine (mut*) is shown in dark blue. The differences between the proposed ABP51975 and ABY76247 vaccines are shown bold in "mut*".

AUTHORS' CONTRIBUTIONS

C. P. Mikita provided the medical significance of the work and researched the early literature. E. A. Padlan did the sequence and structural analyses.

REFERENCES

Caton AJ, Brownlee GG, Yewdell JW, Gerhard W. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 Subtype). Cell 1982; 31(2 Pt 1):417-27.

Connolly ML. Analytical molecular surface calculation. J Appl Crystallogr 1983; 16:548-558.

De Genst E, Areskoug D, Decanniere K, Muyldermans S, Andersson K. Kinetic and affinity predictions of a protein-protein interaction using multivariate experimental design. J Biol Chem 2002; 277:29897-29907.

Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, Vasisht N, Steinhauer DA, Daniels RS, Elliot A, Wiley DC, Skehel JJ. The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science 2004; 303:1838-1842.

Grantham R. Amino acid difference formula to help explain protein evolution. Science 1974; 185(4154):862-864.

Ha Y, Stevens DJ, Skehel JJ, Wiley DC. X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. Proc Natl Acad Sci USA 2001; 98(20):11181-11186.

Hutchinson EG, Thornton JM. A revised set of potentials for beta-turn formation in proteins. Protein Science 1994; 3(12):2207-2216.

Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: exploring protein sequences for globularity and disorder. Nucleic Acid Res 2003; 31(13):3701-3708.

Mikita CP, Padlan EA. Design of possibly universal vaccines against seasonal influenza. Phil Sci Letts 2021; 14(1):12-28.

Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. Biophys J 1998; 75(1):422-427.

Padlan EA. A novel method for designing vaccines against constantly mutating pathogens. Phil J Sci 2008; 137:39-51.

Padlan EA. A method for designing molecules for use in directing the antibody response to a chosen region of a protein antigen. Phil Sci Letts 2010; 3(2):36-47.

Peitsch MC. Protein modeling by E-mail Bio/Technology 1995; 13:658-660.

Sandberg M, Eriksson L, Jonsson J, Sjostrom M, Wold S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem 1998; 41(14):2481-2491.

Saelens X. Tailor-made cleavage site attenuates influenza B. Expert Rev Vaccines 2012; 11(2):159-161.

Sauter NK, Hanson JE, Glick GD, Brown JH, Crowther RL, Park SJ, Skehel JJ, Wiley DC. Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and x-ray crystallography. Biochemistry 1992; 31(40):9609–9621.

Sheriff S, Hendrickson WA, Stenkamp RE, Sieker LC, Jensen LH. Influence of solvent accessibility and intermolecular contacts on atomic mobilities in hemerythrin. Proc Natl Acad Sci USA 1985; 82(4):1104-1107.

Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM. Mapping the antigenic and genetic evolution of influenza virus. Science 2004; 305 (5682):371-376

Stech J, Garn H, Herwig A, Stech O, Dauber B, Wolff T, Mettenleiter TC, Klenk H-D. Influenza B Virus With Modified Hemagglutinin Cleavage Site as a Novel Attenuated Live Vaccine. J Infect Dis 2011; 204(10):1483-1490.

Stray SJ, Pittman LB. Subtype and antigenic site-specific differences in biophysical influences on evolution of influenza virus hemagglutinin. Virol J 2012; 9:91-107.

Street AG, Mayo SL. Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. Proc Natl Acad Sci USA 1999; 96(16):9074-9076.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018; 46(W1):W296-W303.

Wilson IA, Skehel JJ, Wiley DC. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 1981; 289(5796):366–373.