Design of possibly universal vaccines against seasonal influenza

Cecilia P. Mikita[1] and 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

The design of possibly universal vaccines against seasonal influenza is proposed. The vaccines are based on the hemagglutinin glycoprotein and are designed to focus the immune response to the cleavage site of the molecule. The chemical reactivity of the variable regions is reduced (de-Antigenized) by judicious amino acid replacements. Structurally important regions are preserved. The proposed vaccines make no distinction between trivalent and quadrivalent seasonal vaccines.

KEYWORDS

yearly influenza vaccine, hemagglutinin, cleavage site epitope, universal vaccine

INRODUCTION

The World Health Organization (WHO) Global Influenza Surveillance and Response System (GISRS) is a worldwide system of public health institutions that monitors and makes recommendations for influenza viruses of public health concern. GISRS currently consists of 146 National Influenza Centers and 122 WHO Member States, 6 WHO Collaborating Centers for Influenza, 4 WHO Essential Regulatory Laboratories, and 13 WHO H5 Reference Laboratories. The Centers for Disease Control and Prevention (CDC) Influenza Division is one of six WHO Collaborating Centers for Influenza and receives and tests thousands of influenza viruses from around the world each year. The GISRS network analyzes the following information when evaluating influenza viruses: surveillance data with epidemiologic and clinical findings, viral antigenic characterization with antigenic cartography, genetic characteristics of circulating influenza viruses as compared to vaccine viruses, human serology studies with influenza virus vaccines, virus fitness forecasting, antiviral resistance, vaccine effectiveness in previous and current influenza seasons, and availability of candidate vaccine viruses. In February of each year, these data and other findings are evaluated by experts from the GISRS, the OIE/FAO Network of expertise on animal influenza, academic institutions, and other national and regional institutions. The WHO publishes the final recommendations for the seasonal influenza vaccine viruses for the Northern Hemisphere (https://www.who.int/influenza/vaccines/virus/recommendations/en/). The CDC then presents this information to an advisory committee of the Food and Drug Administration for the final decision about vaccine viruses to be distributed in the United States for the next influenza season. Separate recommendations for candidate vaccine viruses for egg-based versus cell-based manufacturing are made.

*Corresponding author
Email Address: eduardo.padlan@gmail.com
Date received: October 22, 2020
Date revised: January 25, 2021
Date accepted: January 31, 2021

There are four types of influenza virus: A, B, C, and D. Yearly vaccines are made for only types A and B. Quadrivalent vaccines include two subtypes of influenza A viruses and two lineages of influenza B viruses. Trivalent vaccines include two subunits of influenza A viruses and one type B virus. There are two glycoprotein molecules on the viral surface, namely, hemagglutinin and neuraminidase, which are critical for the virus to infect and reproduce in the target organism. For type A, there are 18 known subtypes of hemagglutinin (named: H1 to H18) and 9 subtypes of neuraminidase (N1 to N9), and a particular type A influenza virus is characterized by the hemagglutinin and neuraminidase combination on its surface, e.g., H3N2, etc. Lately, humans are plagued by three viral types: seasonal H3N2, pandemic H1N1, and seasonal B. The influenza virus is constantly mutating so that new vaccines are proposed and produced for every season. It takes 6-8 months to produce, approve, and distribute manufactured influenza vaccines.

The hemagglutinin molecule on the surface of the virus is used 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 could effectively interfere with the proteolysis. The cleavage site region is the focus of our vaccine design.

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 a possibly universal vaccine against seasonal H3N2, pandemic H1N1, and seasonal B influenza that could be useful for all seasons.

MATERIALS AND METHODS

Sequence data:

We limit our analyses to the hemagglutinin sequences of the WHO-recommended composition of the H3N2, pandemic H1N1, B influenza trivalent vaccines (Northern Hemisphere) for the years 1998 to 2021. The sequences are available from the National Center for Biological Information (NCBI) database (www.ncbi.nlm.nih.gov), or from the Global Initiative on Sharing All Influenza Data (GISAID) (www.gisaid.org).

The hemagglutinin sequences of the WHO-recommended composition of the H3N2 influenza virus trivalent vaccines (Northern Hemisphere) for the years 1998 to 2021 are listed below starting with the database entry code, followed by the strain name, then the year for which it is recommended:

ABQ10136	(A/Sydney/5/1997)	1998-2000
AAT08002	(A/Moscow/10/1999)	2000-2004
ACF54565	(A/Wyoming/03/2003)	2004-2005
ABW80975	(A/California/07/2004)	2005-2006
ABW80978	(A/Wisconsin/67/2005)	2006-2008
ABW23353	(A/Brisbane/10/2007)	2008-2010
ACS71642	(A/Perth/16/2009)	2010-2012
AGB08328	(A/Victoria/361/2011)	2012-2014
AGL06219	(A/Texas/50/2012)	2014-2015
EPI_ISL_230377	(A/Switzerland/9715293/2013)	2015-2016
EPI_ISL_165554	(A/Hong Kong/4801/2014)	2016-2018
EPI_ISL_225834	(A/Singapore/INF-IMH_16_0019/2016)	2018-2019
AVG71503	(A/Kansas/14/2017)	2019-2020
EPI_ISL_412381	(A/Hong Kong/2671/2019)	2020-2021

The hemagglutinin sequences of the WHO-recommended composition of the pandemic H1N1 influenza virus trivalent vaccines (Northern Hemisphere) for the years 1998 to 2021 are:

ACF41867	(A/Beijing/262/1995)	1998-2000
ACF41878	(A/New Caledonia/20/1999)	2000-2007
ABU99109	(A/Solomon Islands/3/2006)	2007-2008
ADE28750	(A/Brisbane/59/2007)	2008-2010
ADE28750	(A/California/07-JRO09/2009)	2010-2017
APB91426	(A/Michigan/45/2015)	2017-2019
EPI_ISL_306335	(A/Brisbane/02/2018)	2019-2020
EPI_ISL_391021	(A/Guangdong-Maonan/SWL1536/2019)	2020-2021

The hemagglutinin sequences of the WHO-recommended composition of the seasonal B influenza virus trivalent vaccines (Northern Hemisphere) for the years 1998 to 2021 are:

ACR15721	(B/Harbin/7/1994)	1998-2000
ABN50503	(B/Yamanashi/166/1998)	2000-2001
AAT69455	(B/Victoria/504/2000)	2001-2002
ABL77145	(B/Hong Kong/1434/2002)	(2002-2004)
ACF54180	(A/Wisconsin/67/2005)	2006-2008
ACO05957	(A/Brisbane/10/2007)	2008-2010
ACA33493	(A/Perth/16/2009)	2010-2012
EPI_ISL_28587	(A/Victoria/361/2011)	2012-2014
AET22022.	(A/Texas/50/2012)	2014-2015
AGL06036.	(A/Switzerland/9715293/2013)	2015-2016
EPI_ISL_168822	(A/Hong Kong/4801/2014)	2016-2018
AFH57909	(A/Singapore/INF-IMH_16_0019/2016)	2018-2019
EPI_ISL_257735	(A/Kansas/14/2017)	2019-2020
QCG86174	(A/Hong Kong/2671/2019)	2020-2021

The WHO-recommended B influenza viruses included in the quadrivalent vaccines are discussed later.

Table 1: Hemagglutinins from WHO-recommended H3N2 component of trivalent influenza vaccines for 1998-2021 (Northern Hemisphere)

```
      10        20        30        40        50        60
        |         |         |         | c  cc   |   e     |
STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICDSPHRILDGENCTLID      1998-2000
STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICDSPHQILDGENCTLID      2000-2004
STATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSSTGGICDSPHQILDGENCTLID      2004-2005
STATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSSTGGICDSPHQILDGENCTLID      2005-2006
STATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSSTGGICDSPHQILDGENCTLID      2006-2008
STATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSSTGEICDSPHQILDGENCTLID      2008-2010
STATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSSTGEICDSPHQILDGKNCTLID      2010-2012
STATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQNSSIGEICDSPHQILDGENCTLID      2012-2014
STATLCLGHHAVPNGTIVKTITNDRIEVTNATELVQNSSIGEICDSPHQILDGENCTLID      2014-2015
STATLCLGHHAVPNGTIVKTITNDRIEVTNATELVQNSSIGEICDSPHQILDGENCTLID      2015-2016
STATLCLGHHAVPNGTIVKTITNDRIEVTNATELVQNSSIGEICDSPHQILDGENCTLID      2016-2018
STATLCLGHHAVPNGTIVKTITNDRIEVTNATELVQNSSIGEICDSPHQILDGENCTLID      2018-2019
STATLCLGHHAVPNGTIVKTITNDRIEVTNATELVQNSSIGEICDSPHQILDGENCTLID      2019-2020
STATLCLGHHAVPNGTIVKTITNDRIEVTNATELVQNSSIGEICDSPHQILDGGNCTLID      2020-2021
                *       *           *  * *      *    *
       70        80        90        100       110       120
      e  |   ee    |         +   d     |         |   a a   |
ALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNESFNWT      1998-2000
ALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNESFNWT      2000-2004
ALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNESFNWA      2004-2005
ALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNESFNWT      2005-2006
ALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNDESFNWT      2006-2008
ALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNESFNWT      2008-2010
ALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNESFNWT      2010-2012
ALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNESFNWT      2012-2014
ALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNESFNWN      2014-2015
ALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNESFNWA      2015-2016
ALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNESFNWT      2016-2018
ALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFKNESFNWT      2018-2019
ALLGDPQCDGFQNKKWDLFVERNKAYSNCYPYDVPDYASLRSLVASSGTLEFNNESFNWA      2019-2020
ALLGDPQCDGFQNKKWDLFVERSRAYSNCYPYDVPDYASLRSLVASSGTLEFKNESFNWA      2020-2021
      *       *       **                                **     *
       130       140       150       160       170       180
  a a   a|    aaaa |    + bb b b   b   |   d d   |    +    b
GVAQNGTSYACKRSSIKSFFSRLNWLHQLKYKYPALNVTMPNNDKFDKLYIWGVHHPSTD      1998-2000
GVAQNGTSSSCKRRSIKSFFSRLNWLHQLKYRYPALNVTMPNNDKFDKLYIWGVHHPSTD      2000-2004
GVTQNGTSSACKRRSNKSFFSRLNWLTHLKYKYPALNVTMPNNEKFDKLYIWGVHHPVTD      2004-2005
GVTQNGTSSSCKRRSNNSFFSRLNWLTHLKFKYPALNVTMPNNEKFDKLYIWGVHHPGTN      2005-2006
GVTQNGTSSSCKRRSNNSFFSRLNWLTHLKFKYPALNVTMPNNEKFDKLYIWGVHHPVTD      2006-2008
GVTQNGTSSACIRRSNNSFFSRLNWLTHLKFKYPALNVTMPNNEKFDKLYIWGVHHPGTD      2008-2010
GVTQNGTSSACIRRSKNSFFSRLNWLTHLNFKYPALNVTMPNNEQFDKLYIWGVHHPGTD      2010-2012
GVTQNGTSSACIRRSNNSFFSRLNWLTQLNFKYPALNVTMPNNEQFDKLYIWGVHHPVTD      2012-2014
GVTQNGTSSACIRRSNNSFFSRLNWLTHLNFKYPALNVTMPNNEQFDKLYIWGVHHPVTD      2014-2015
GVTQNGTSSSCIRGSNSSFFSRLNWLTHLNSKYPALNVTMPNNEQFDKLYIWGVHHPGTD      2015-2016
GVTQNGTSSACIRRSSSSFFSRLNWLTHLNYTYPALNVTMPNNEQFDKLYIWGVHHPGTD      2016-2018
GVTQNGTSSACIRGSSSSFFSRLNWLTHLNYTYPALNVTMPNKEQFDKLYIWGVHHPGTD      2018-2019
GVTQNGTSSSCIRGSKSSFFSRLNWLTHLNSKYPALNVTMPNNEQFDKLYIWGVHHPGTD      2019-2020
GVTQNGKSFSCIRGSSSSFFSRLNWLTHLNYIYPALNVTMPNKEQFDKLYIWGVHHPVTD      2020-2021
  *   * ** * * **          ** ***          ***            * *
       190       200       210       220       230       240
bb  b+ bb|  d     d|    d   d|       + + d       |     d   |
SDQTSIYAQASGRVTVSTKRSQQTVIPNIGSRPWVRGISSRISIHWTIVKPGDILLINST      1998-2000
SDQTSLYTQASGRVTVSTKRSQQTVIPNIGSRPWVRGISSRISIYWTIVKPGDILLINST      2000-2004
SDQISLYAQASGRITVSTKRSQQTVIPNIGYRPRVRDISSRISIYWTIVKPGDILLINST      2004-2005
NDQISLYTQASGRITVSTKRSQQTVIPNIGSRPRVRDIPSRISIYWTIVKPGDILLINST      2005-2006
NDQIFLYAQASGRITVSTKRSQQTVIPNIGSRPRIRNIPSRISIYWTIVKPGDILLINST      2006-2008
NDQIFPYAQASGRITVSTKRSQQTVIPNIGSRPRVRNIPSRISIYWTIVKPGDILLINST      2008-2010
KDQIFLYAQASGRITVSTKRSQQTVSPNIGSRPRVRNIPSRISIYWTIVKPGDILLINST      2010-2012
KDQIFLYAQSSGRITVSTKRSQQAVIPNIGYRPRIRNIPSRISIYWTIVKPGDILLINST      2012-2014
KDQIFLYAQPSGRITVSTKRSQQAVIPNIGFRPRIRNIPSRISIYWTIVKPGDILLINST      2014-2015
KDQIFLYAQSSGRITVSTKRSQQAVIPNIGSRPRIRDIPSRISIYWTIVKPGDILLINST      2015-2016
```

```
KDQIFLYAQSSGRITVSTKRSQQAVIPNIGSRPRIRDIPSRISIYWTIVKPGDILLINST      2016-2018
KDQIFLYAQSSGRITVSTKRSQQAVIPNIGSRPRIRDIPSRISIYWTIVKPGDILLINST      2018-2019
KDQISLYAQSSGRITVSTKRSQQAVIPNIGSRPRIRDIPSRISIYWTIVKPGDILLINST      2019-2020
KDQISLYAQSSGRITVSTKRSQQAVIPNIGSRPRIRNIPSRISIYWTIVKPGDILLINST      2020-2021
*  *** * *   *         * *    *  ** * *     *
        250       260       270       280       290       300
         | e e     |       c c         |         |         |
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCNSECITPNGSIPNDKPFQNVNRITYGACPRY      1998-2000
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCNSECITPNGSIPNDKPFQNVNRITYGACPRY      2000-2004
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCNSECITPNGSIPNDKPFQNVNRITYGACPRY      2004-2005
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCNSECITPNGSIPNDKPFQNVNRITYGACPRY      2005-2006
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCNSECITPNGSIPNDKPFQNVNRITYGACPRY      2006-2008
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCNSECITPNGSIPNDKPFQNVNRITYGACPRY      2008-2010
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCNSECITPNGSIPNDKPFQNVNRITYGACPRY      2010-2012
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCNSECITPNGSIPNDKPFQNVNRITYGACPRY      2012-2014
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPFQNVNRITYGACPRY      2014-2015
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPFQNVNRITYGACPRY      2015-2016
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPFQNVNRITYGACPRY      2016-2018
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPFQNVNRITYGACPRY      2018-2019
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPFQNVNRITYGACPRY      2019-2020
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPFQNVNRITYGACPRY      2020-2021
                            *
        310       320       330       340       350       360
         |         |         |         |         |         |
VKQNTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGTGQAADLK      1998-2000
VKQNTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMMDGWYGFRHQNSEGTGQAADLK      2000-2004
VKQNTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGTGQAADLK      2004-2005
VKQNTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGIGQAADLK      2005-2006
VKQNTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGIGQAADLK      2006-2008
VKQNTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGIGQAADLK      2008-2010
VKQNTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLK      2010-2012
VKQSTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLK      2012-2014
VKQSTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLK      2014-2015
VKQSTLKLATGMRNVPERQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLK      2015-2016
VKHSTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLK      2016-2018
VKHSTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLK      2018-2019
VKQSTLKLATGMRNVPERQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLK      2019-2020
VKQSTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLK      2020-2021
  **          *                       *             *
        370       380       390       400       410       420
         |         |         |         |         |         |
STQAAINQINGKLNRLIEKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      1998-2000
STQAAINQINGKLNRLIEKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      2000-2004
STQAAINQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      2004-2005
STQAAINQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      2005-2006
STQAAINQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      2006-2008
STQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      2008-2010
STQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      2010-2012
STQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      2012-2014
STQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      2014-2015
STQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      2015-2016
STQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      2016-2018
STQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRVQDLEKYVEDTKIDLWSYNAELL      2018-2019
STQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELL      2019-2020
STQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRVQDLEKYVEDTKIDLWSYNAELL      2020-2021
      *           *                     *
        430       440       450       460       470       480
         |         |         |         |         |         |
VALENQHTIDLTDSEMNKLFERTRKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDH      1998-2000
VALENQHTIDLTDSEMNKLFERTRKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDH      2000-2004
VALENQHTIDLTDSEMNKLFERTKKQLRENAEDMGNGCFKIYHKCDNACIESIRNGTYDH      2004-2005
VALENQHTIDLTDSEMNKLFERTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDH      2005-2006
VALENQHTIDLTDSEMNKLFERTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDH      2006-2008
```

```
VALENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDH      2008-2010
VALENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDH      2010-2012
VALENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDH      2012-2014
VALENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDH      2014-2015
VALENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDH      2015-2016
VALENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDH      2016-2018
VALENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIESIRNETYDH      2018-2019
VALENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACMGSIRNGTYDH      2019-2020
VALENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNETYDH      2020-2021
                     * *                          **    *

        490
         |
DVYRDEALNNRFQI                                                     1998-2000
DVYRDEALNNRFQI                                                     2000-2004
DVYRDEALNNRFQI                                                     2004-2005
DVYRDEALNNRFQI                                                     2005-2006
DVYRDEALNNRFQI                                                     2006-2008
DVYRDEALNNRFQI                                                     2008-2010
DVYRDEALNNRFQI                                                     2010-2012
DVYRDEALNNRFQI                                                     2012-2014
DVYRDEALNNRFQI                                                     2014-2015
DVYRDEALNNRFQI                                                     2015-2016
NVYRDEALNNRFQI                                                     2016-2018
NVYRDEALNNRFQI                                                     2018-2019
NVYRDEALNNRFQI                                                     2019-2020
NVYRDEALNNRFQI                                                     2020-2021
*
```

Positions where any change had occurred from year to year are indicated by asterisks below the sequences. The residues, whose alpha-carbons are within 17A of that of the cleavage site (Arg321), are shown bold and doubly underlined in the sequence of GISAID Entry EPI_ISL_412381 (A/Hong Kong/2671/2019) which is the (2020-2021) WHO-recommended entry. Possible N-glycosylation sites (N-X-S/T, where X is an amino acid other than proline) are underlined. Residues in the H3 receptor-binding site (Tyr90, Trp145, Thr147, His175, D182, Leu186, Thr218, and Ser220) are from Sauter et al. (1992) (PDB Entry 1HGH) and are indicated by plus signs (+) in red. The H3 antigenic sites from Smith et al. (2004) and the residues in the site are identified by the letters a, b, c, d, and e above them for the sites A, B, C, D, and E, respectively. When a receptor-binding site residue coincides with an antigenic-site residue, in particular residues at positions 147 and 182, the antigenic-site residue is shown in red. Positions where any change had occurred from year to year in the period 1998-2021 are indicated by asterisks (*) below the sequences.

Table 2: Hemagglutinins from WHO-recommended H1N1 component of trivalent influenza vaccines for 1998-2021 (Northern Hemisphere)

```
        10        20        30        40        50        60
         |         |         |         |         |         |
DTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCLLKGIAPLQLGNCSVAGW      1998-2000
DTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCLLKGIAPLQLGNCSVAGW      2000-2007
DTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCLLKGIAPLQLGNCSVAGW      2007-2008
DTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLENSHNGKLCLLKGIAPLQLGNCSVAGW      2008-2010
DTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLRGVAPLHLGKCNIAGW      2010-2017
DTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLRGVAPLHLGKCNIAGW      2017-2019
DTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLGGVAPLHLGKCNIAGW      2019-2020
DTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLRGVAPLHLGKCNIAGW      2020-2021
  *                            **      * * *   *  * **

        70        80        90        100       110       120
         ee eee    |         |+        |         |    e    |
ILGNPECESLISKESWSYIVETPNPENGTCYPGYFADYEELREQLSSVSSFERFEIFPKE      1998-2000
ILGNPECELLISKESWSYIVETPNPENGTCYPGYFADYEELREQLSSVSSFERFEIFPKE      2000-2007
ILGNPECELLISRESWSYIVEKPNPENGTCYPGHFADYEELREQLSSVSSFERFEIFPKE      2007-2008
ILGNPECELLISKESWSYIVEKPNPENGTCYPGHFADYEELREQLSSVSSFERFEIFPKE      2008-2010
ILGNPECESLSTASSWSYIVETPSSDNGTCYPGDFIDYEELREQLSSVSSFERFEIFPKT      2010-2017
ILGNPECESLSTASSWSYIVETSNSDNGTCYPGDFINYEELREQLSSVSSFERFEIFPKT      2017-2019
ILGNPECESLSTARSWSYIVETSNSDNGTCYPGDFINYEELREQLSSVSSFERFEIFPKT      2019-2020
ILGNPECESLSTARSWSYIVETSNSDNGTCYPGDFINYEELREQLSSVSSFERFEIFPKT      2020-2021
        * ****       *****       * **                    *

        130       140       150       160       170       180
   aa    |      d  d d       + +baab  aa aaa c   c        c+
SSWPNH-TVTGVTASCSHNGKSSFYRNLLWLTEKNGLYPNLSNSYVNNKEKEVLVLWGVH      1998-2000
SSWPNH-TVTGVSASCSHNGKSSFYRNLLWLTGKNGLYPNLSKSYVNNKEKEVLVLWGVH      2000-2007
SSWPNH-TTTGVSASCSHNGESSFYKNLLWLTGKNGLYPNLSKSYANNKEKEVLVLWGVH      2007-2008
SSWPNH-TVTGVSASCSHNGESSFYRNLLWLTGKNGLYPNLSKSYANNKEKEVLVLWGVH      2008-2010
```

```
SSWPNHDSNKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLSKSYINDKGKEVLVLWGIH      2010-2017
SSWPNHDSNKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLNQSYINDKGKEVLVLWGIH      2017-2019
SSWPNHDSNKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLNQTYINDKGKEVLVLWGIH      2019-2020
SSWPNHDSDKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLNQTYINDKGKEVLVLWGIH      2020-2021
      ****  * * * * **   *  *  ** ***  * *** * * *        *
        190       200       210       220       230       240
 +   +  bb+ b b    |   c     |          |dd+       |      c  |
HPSNIRDQRAIYHTENAYVSVVSSHYSRRFTPEIAKRPKVRGQEGRINYYWTLLEPGDTI      1998-2000
HPPNIGNQRALYHTENAYVSVVSSHYSRRFTPEIAKRPKVRDQEGRINYYWTLLEPGDTI      2000-2007
HPPNIGDQRALYHKENAYVSVVSSHYSRKFTPEIAKRPKVRDQEGRINYYWTLLEPGDTI      2007-2008
HPPNIGNQKALYHTENAYVSVVSSHYSRKFTPEIAKRPKVRDQEGRINYYWTLLEPGDTI      2008-2010
HPSTSADQQSLYQNADAYVFVGSSRYSKKFKPEIAIRPKVRDREGRMNYYWTLVEPGDKI      2010-2017
HPSTTADQQSLYQNADAYVFVGTSRYSKKFKPEIATRPKVRDQEGRMNYYWTLVEPGDKI      2017-2019
HPPTTADQQSLYQNADAYVFVGTSRYSKKFKPEIATRPKVRDREGRMNYYWTLVEPGDKI      2019-2020
HPPTIAVQESLYQNADAYVFVGTSRYSKKFKPEIATRPKVRDQEGRMNYYWTLVEPGDKI      2020-2021
  ***** *** ****   * ** *  ** *    *     **   *      *    *
        250       260       270       280       290       300
         |         |         |c        |         |         |
IFEANGNLIAPWYAFALSRGFGSGIITSNAPMNECDAKCQTPQGAINSSLPFQNVHPVTI      1998-2000
IFEANGNLIAPWYAFALSRGFGSGIITSNAPMDECDAKCQTPQGAINSSLPFQNVHPVTI      2000-2007
IFEANGNLIAPRYAFALSRGFGSGIINSNAPMDECDAKCQTPQGAINSSLPFQNVHPVTI      2007-2008
IFEANGNLIAPRYAFALSRGFGSGIINSNAPMDKCDAKCQTPQGAINSSLPFQNVHPVTI      2008-2010
TFEATGNLVVPRYAFAMERNAGSGIIISDTPVHDCNTTCQTPKGAINTSLPFQNIHPITI      2010-2017
TFEATGNLVVPRYAFTMERNAGSGIIISDTPVHDCNTTCQTPEGAINTSLPFQNIHPITI      2017-2019
TFEATGNLVVPRYAFTMERNAGSGIIISDTPVHDCNTTCQTAEGAINTSLPFQNVHPVTI      2019-2020
TFEATGNLVVPRYAFTMERDAGSGIIISDTPVHDCNTTCQTPEGAINTSLPFQNVHPITI      2020-2021
*   *   ** *   *** **     * ** *** ***   **    *      *  *
        310       320       330       340       350       360
         |         |         |         |         |         |
GECPKYVRSTKLRMVTGLRNIPSIQSRGLFGAIAGFIEGGWTGMMDGWYGYHHQNEQGSG      1998-2000
GECPKYVRSAKLRMVTGLRNIPSIQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSG      2000-2007
GECPKYVRSAKLRMVTGLRNIPSIQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSG      2007-2008
GECPKYVRSAKLRMVTGLRNIPSIQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSG      2008-2010
GKCPKYVKSTKLRLATGLRNIPSIQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSG      2010-2017
GKCPKYVKSTKLRLATGLRNVPSIQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSG      2017-2019
GKCPKYVKSTKLRLATGLRNVPSIQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSG      2019-2020
GKCPKYVKSTKLRLATGLRNVPSIQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSG      2020-2021
 *     * *   **     *                          *
        370       380       390       400       410       420
         |         |         |         |         |         |
YAADQKSTQNAINGITNKVNSVIEKMNTQFTAVGKEFNKLERRMENLNKKVDDGFLDIWT      1998-2000
YAADQKSTQNAINGITNKVNSVIEKMNTQFTAVGKEFNKLERRMENLNKKVDDGFLDIWT      2000-2007
YAADQKSTQNAINGITNKVNSVIEKMNTQFTAVGKEFNKLERRMENLNKKVDDGFIDIWT      2007-2008
YAADQKSTQNAINGITNKVNSVIEKMNTQFTAVGKEFNKLERRMENLNKKVDDGFIDIWT      2008-2010
YAADLKSTQNAIDEITNKVNSVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWT      2010-2017
YAADLKSTQNAIDKITNKVNSVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWT      2017-2019
YAADLKSTQNAIDKITNKVNSVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWT      2019-2020
YAADLKSTQNAIDKITNKVNSVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWT      2020-2021
    *       **                          *  * *           *
        430       440       450       460       470       480
         |         |         |         |         |         |
YNAELLVLLENERTLDFHDSNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCNNECMESVK      1998-2000
YNAELLVLLENERTLDFHDSNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCNNECMESVK      2000-2007
YNAELLVLLENERTLDFHDSNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCNDECMESVK      2007-2008
YNAELLVLLENERTLDFHDSNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCNDECMESVK      2008-2010
YNAELLVLLENERTLDYHDSNVKNLYEKVRSQLKNNAKEIGNGCFEFYHKCDNTCMESVK      2010-2017
YNAELLVLLENERTLDYHDSNVKNLYEKVRNQLKNNAKEIGNGCFEFYHKCDNTCMESVK      2017-2019
YNAELLVLLENERTLDYHDSNVKNLYEKVRNQLKNNAKEIGNGCFEFYHKCDNTCMESVK      2019-2020
YNAELLVLLENERTLDYHDSNVKNLYEKVRNQLKNNAKEIGNGCFEFYHKCDNTCMESVK      2020-2021
               *           **                    ***
        490
         |
NGTYDYPKYSEESKLNR                                                  1998-2000
```

```
NGTYDYPKYSEESKLNR                                                    2000-2007
NGTYDYPKYSEESKLNR                                                    2007-2008
NGTYDYPKYSEESKLNR                                                    2008-2010
NGTYDYPKYSEEAKLNR                                                    2010-2017
NGTYDYPKYSEEAKLNR                                                    2017-2019
NGTYDYPKYSEEAKLNR                                                    2019-2020
NGTYDYPKYSEEAKLNR                                                    2020-2021
            *
```

Positions where any change had occurred from year to year are indicated by asterisks below the sequences. The residues, whose alpha-carbons are within 17A of that of the cleavage site (Arg327), are shown bold and doubly underlined in the sequence of GISAID Entry EPI_ISL_391021 (A/Guangdong-Maonan/SWL1536/2019) which is the (2020-2021) WHO-recommended entry. Possible N-glycosylation sites (N-X-S/T, where X is an amino acid other than proline) are underlined. Residues in the H1 receptor-binding site (Tyr91, Trp150, Val152, His180, Pro183, Trp187, Leu198, and Gln223) are from Gamblin et al. (2004) (PDB Entry 1PVZ) and are indicated by plus signs (+) in red. Positions where any change had occurred from year to year in the period 1998-2021 are indicated by asterisks (*) below the sequences.

Table 3: Hemagglutinins from WHO-recommended B component of trivalent influenza vaccines for 1998-2021 (Northern Hemisphere)

```
        10        20        30        40        50        60
         |         |         |         |      cc |     e e |
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSHFANLKGTKTRGKLCPNCLNC      1998-2000
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSHFANLKGTKTRGKLCPTCLNC      2000-2001
DRICTGITSSNSPHVVKTATQGEVNVTGAIPLTTTPTKSHFANLKGTKTRGKLCPTCLNC      2001-2002
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSHFANLKGTKTRGKLCPKCLNC      2002-2004
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSYFANLKGTRTRGKLCPDCLNC      2004-2006
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSHFANLKGTETRGKLCPKCLNC      2006-2008
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSYFANLKGTRTRGKLCPDCLNC      2008-2009
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSHFANLKGTETRGKLCPKCLNC      2009-2012

DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSYFANLKGTRTRGKLCPDCLNC      2012-2013
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSYFANLKGTKTRGKLCPDCLNC      2013-2015
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSYFANLKGTRTRGKLCPDCLNC      2015-2016
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSHFANLKGTETRGKLCPKCLNC      2016-2018
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSHFANLKGTETRGKLCPKCLNC      2018-2020
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSHFANLKGTETRGKLCPKCLNC      2020-2021
                            *         *       *       *

        70        80        90       100       110       120
       ee|e e ee   cc         |    +    |         |     c   |
TDLDVALGRPMCVGTTPSAKASILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYENIRLS      1998-2000
TDLDVALGRPMCVGVTPSAKASILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYEKIRLS      2000-2001
TDLDVALGRPMCVGITPSAKASILHEVKPVTSGCFPIMHDRTKIRQLPNLLRGYEKIRLS      2001-2002
TDLDVALGRPKCTGNIPSAKVSILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYERIRLS      2002-2004
TDLDVALGRPMCVGTTPSAKASILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYENIRLS      2004-2006
TDLDVALGRPKCTGNIPSARVSILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYEHIRLS      2006-2008
TDLDVALGRPMCVGTTPSAKASILHEVKPVTSGCFPIMHDRTKIRQLPNLLRGYENIRLS      2008-2009
TDLDVALGRPKCTGKIPSARVSILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYEHIRLS      2009-2012

TDLDVALGRPMCVGTTPSAKASILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYENIRLS      2012-2013
TDLDVALGRPMCVGTTPSAKASILHEVRPVTSGCFPIMHDRTKIRQLANLLRGYENIRLS      2013-2015
TDLDVALGRPMCVGTTPSAKASILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYEKIRLS      2015-2016
TDLDVALGRPKCTGKIPSARVSILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYEHIRLS      2016-2018
TDLDVALGRPKCTGKIPSARVSILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYEHVRLS      2018-2020
TDLDVALGRPKCTGKIPSARVSILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYEHVRLS      2020-2021
          * * **   **      *                    *       **

        130       140       150       160       170       180
dd  dddd |     aa  |a    aaaaa   a   + |   bbbbbb |         |
TQNVINAEKAPGGPYRLGTSGSCPNATSRSGFFATMAWAVPR-DDNKTATNPLTVEVPYVC      1998-2000
TQNVINAEKAPGGPYRLGTSGSCPNATSRSGFFATMAWAVPK-DNNKTATNPLTVEVPHIC      2000-2001
TQNVINAEKAPGGPYRLGTSGSCPNATSKSGFFATMAWAVPM-DNNKTATNPLTVEVPHIC      2001-2002
NHNVINAEEAPGGPYKIGTSGSCPNVTNGNGFFATMAWAVPKNENNKTATNSLTIEVPYIC      2002-2004
TQNVIDAENAPGGPYRLGTSGSCPNATSKSGFFATMAWAVPK-DNNKNATNPLTVEVPYVC      2004-2006
THNVINAENAPGGPYKIGTSGSCPNVTNGNGFFATMAWAVPKNDNNKTATNSLTIEVPYIC      2006-2008
TQNVIDAEKAPGGPYRLGTSGSCPNATSKSGFFATMAWAVPK-DNNKNATNPLTVEVPYIC      2008-2009
THNVINAENAPGGPYKIGTSGSCPNITNGNGFFATMAWAVPKNDKNKTATNPLTIEVPYIC      2009-2012
```

```
TQNVIDAEKAPGGPYRLGTSGSCPNATSKIGFFATMAWAVPK-DNYKNATNPLTVEVPYIC     2012-2013
TQNVIDAEKAPGGPYRLGTSGSCPNATSKSGFFATMAWAVPK-DNNKNATNPLTVEVPYIC     2013-2015
TQNVIDAEKAPGGPYRLGTSGSCPNATSKIGFFATMAWAVPK-DNYKNATNPLTVEVPYIC     2015-2016
THNVINAENAPGGPYKIGTSGSCPNITNGNGFFATMAWAVPKNDKNKTATNPLTIEVPYIC     2016-2018
THNVINAEGAPGGPYKIGTSGSCPNITNGNGFFATMAWAVP--DKNKTATNPLTIEVPYVC     2018-2020
THNVINAEDAPGRPYEIGTSGSCPNITNGNGFFATMAWAVPK---NKTATNPLTIEVPYIC     2020-2021
**   *  *   *  **        * ***           ***** *   *  *   **

        190       200       210       220       230       240
 ddd     |  +   bbbb bb+   bb|         |         |         |
TEGEDQITVWGFHSDNKAQMKNLYGDSNPQKFTSSANGVTTHYVSQIGGFPDQTEDGGLP      1998-2000
TKEEDQITVWGFHSDDKTQMKNLYGDSNPQKFTSSANGVTTHYVSQIGGFPDQTEDGGLP      2000-2001
TKEEDQITVWGFHSDYKTQMKNLYGDSNPQKFTSSANGITTHYVSQIGGFPEQTEDGGLP      2001-2002
TEGEDQITVWGFHSDSETQMAKLYGDSKPQKFTSSANGVTTHYVSQIGGFPNQTEDGGLP      2002-2004
TEGEDQITVWGFHSDNKTPMKNLYGDSNPQKFTSSANGVTTHYVSQIGGFPAQTEDGGLP      2004-2006
TEGEDQITVWGFHSDNEIQMAKLYGDSKPQKFTSSANGVTTHYVSQIGGFPNQTEDGGLP      2006-2008
TEGEDQITVWGFHSDDKTQMKNLYGDSNPQKFTSSANGVTTHYVSQIGSFPDQTEDGGLP      2008-2009
TEGEDQITVWGFHSDNETQMAKLYGDSKPQKFTSSANGVTTHYVSQIGGFPNQTEDGGLP      2009-2012

TEGEDQITVWGFHSDNKTQMKSLYGDSNPQKFTSSANGVTTHYVSQIGDFPDQTEDGGLP      2012-2013
AEGEDQITVWGFHSDDKTQMKNLYGDSNPQKFTSSANGVTTHYVSQIGGFPDQTEDGGLP      2013-2015
TEGEDQITVWGFHSDDKTQMKSLYGDSNPQKFTSSANGVTTHYVSQIGDFPDQTEDGGLP      2015-2016
TEGEDQITVWGFHSDDETQMAKLYGDSKPQKFTSSANGVTTHYVSQIGGFPNQTEDGGLP      2016-2018
TEGEDQITVWGFHSDNETQMAKLYGDSKPQKFTSSANGVTTHYVSQIGGFPNQTEDGGLP      2018-2020
TEGEDQITVWGFHSDXETQMAKLYGDSKPQKFTSSANGVTTHYVSQIGGFPNQTEDGGLP      2020-2021
***             **** **     *          *          *  *

        250       260       270       280       290       300
         |dd  dddd |         |         c|   c     |         |
QSGRIVVDYMVQKPGKTGTIVYQRGVLLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      1998-2000
QSGRIVVDYMVQKPGKTGTIVYQRGILLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2000-2001
QSGRIVVDYMVQKPGKTGTIVYQRGILLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2001-2002
QSGRIVVDYMVQKSGKTGTITYQRGILLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2002-2004
QSGRIVVDYMVQKPRKTGTIVYQRGVLLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2004-2006
QSGRIVVDYMVQKSGKTGTITYQRGILLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2006-2008
QSGRIVVDYMMQKPGKTGTIVYQRGVLLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2008-2009
QSGRIVVDYMVQKSGKTGTITYQRGILLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2009-2012

QSGRIVVDYMMQKPGKTGTIVYQRGVLLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2012-2013
QSGRIVVDYMMQKPGKTGTIVYQRGVLLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2013-2015
QSGRIVVDYMMQKPGKTGTIVYQRGVLLPQKVWCASGRSKVIKGSLPLIGEADCLHEEYG      2015-2016
QSGRIVVDYMVQKSGKTGTITYQRGILLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2016-2018
QSGRIVVDYMVQKSGKTGTITYQRGILLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2018-2020
QSGRIVVDYMVQKSGKTGTITYQRGILLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG      2020-2021
          *  **     *    *                                  *

        310       320       330       340       350       360
         |         |         |         |         |         |
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      1998-2000
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2000-2001
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2001-2002
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2002-2004
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2004-2006
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2006-2008
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2008-2009
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2009-2012

GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2012-2013
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2013-2015
GLNKSKPYYTGKHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2015-2016
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2016-2018
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2018-2020
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW      2020-2021
```

```
          *

         370        380        390        400        410        420
          |          |          |          |          |          |
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      1998-2000
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2000-2001
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2001-2002
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2002-2004
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2004-2006
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2006-2008
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2008-2009
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2009-2012

EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2012-2013
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2013-2015
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2015-2016
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2016-2018
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2018-2020
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      2020-2021


         430        440        450        460        470        480
          |          |          |          |          |          |
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVDIG      1998-2000
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVDIG      2000-2001
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVDIG      2001-2002
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVEIG      2002-2004
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVDIG      2004-2006
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVEIG      2006-2008
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVEIG      2008-2009
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVEIG      2009-2012

NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVDIG      2012-2013
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVDIG      2013-2015
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVDIG      2015-2016
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVEIG      2016-2018
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVEIG      2018-2020
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVEIG      2020-2021
                                                              *
         490        500        510
          |          |          |
NGCFETKHKCNQTCLDRIAAGTFNAGEFSLPTFDSLNI                            1998-2000
NGCFETKHKCNQTCLDRIAAGTFNAGEFSLPTFDSLNI                            2000-2001
NGCFETKHKCNQTCLDRIAAGTFNAGEFSLPTFDSLNI                            2001-2002
NGCFETKHKCNQTCLDRIAAGTFNAGEFSLPTFDSLNI                            2002-2004
NGCFETKHKCNQTCLDRIAAGTFNAGEFSLPTFDSLNI                            2004-2006
NGCFETKHKCNQTCLDRIAAGTFDAGEFSLPTFDSLNI                            2006-2008
NGCFETKHKCNQTCLDRIAAGTFNAGEFSLPTFDSLNI                            2008-2009
NGCFETKHKCNQTCLDRIAAGTFDAGEFSLPTFDSLNI                            2009-2012

NGCFETKHKCNQTCLDRIAAGTFNAGEFSLPTFDSLNI                            2012-2013
NGCFETKHKCNQTCLDRIAAGTFNAGEFSLPTFDSLNI                            2013-2015
NGCFETKHKCNQTCLDRIAAGTFNAGEFSLPTFDSLNI                            2015-2016
NGCFETKHKCNQTCLDRIAAGTFDAGEFSLPTFDSLNI                            2016-2018
NGCFETKHKCNQTCLDKIAAGTFDAGEFSLPTFDSLNI                            2018-2020
NGCFETKHKCNQTCLDRIAAGTFDAGEFSLPTFDSLNI                            2020-2021
                 *       *
```

Positions where any change had occurred from year to year are indicated by asterisks below the sequences. The residues, whose alpha-carbons are within 17A of that of the cleavage site (Arg346), are shown bold and doubly underlined in the sequence of GISAID Entry EPI_ISL_257735 (B/Colorado/06/2017) which is the (2018-2020) WHO- recommended entry. Possible N-glycosylation sites (N-X-S/T, where X is an amino acid other than proline) are underlined. The B antigenic sites, shown bold and underlined, are from Stray and Pittman (2012). Residues in the antigenic sites are identified by the letters a, b, c, d, and e above them for sites BA, BB1+BB2, BC, BD, and BE of Stray and Pittman (2012), respectively. The location of the antigenic residues beyond position 170 differs by one from those reported by Stray and Pittman (2012) because of the variation in the insertions at positions 161-165. Residues in the B receptor-binding site (Phe95, Trp158, His191, and Tyr202) are from Wang et al. (2007) (PDB Entry 2RFU) and are indicated by plus signs (+) in red. Positions where any change had occurred from year to year in the period 1998-2021 are indicated by asterisks (*) below the sequences.

Sequence differences:
The sequences are listed in Table 1 for seasonal H3N2, Table 2 for pandemic H1N1, and Table 3 for seasonal B influenza. Sequence differences from year to year are indicated in the tables. The differences are predominantly in the HA1 portion (residues 1- 320 in hemagglutinin H3, 1- 326 in H1, and 1- 345 in B). Insertions and deletions are also seen in the listing of the H1 and B sequences (Tables 2 and 3). Indicated in the tables are the positions of the residues in the antigenic sites which have been compiled by Smith et al. (2004) for H3, by Caton et al. (1982) for H1, and by Stray and Pittman (2012) for B hemagglutinins. The receptor-binding sites deduced from x-ray crystallographic structures by Sauter et al. (1992) (PDB Entry 1HGH) for H3, by (Gamblin et al. 2004) (PDB Entry 1RVZ) for H1, and by Ha et al. (2001) (PDB Entry 1JSN) for B hemagglutinin, are also indicated in the tables. Differences in the location of the antigenic site and receptor-binding positions indicated in Tables 1 - 3 and those mentioned in the original references are due to insertions or deletions. Note the changes in the identity of the receptor-binding residues through the years; a structural explanation for these sequence changes is beyond the scope of this study.

Modeling of the hemagglutinin trimers:
Hemagglutinin exists as a trimer on the surface of the virus. Since the process of de-Antigenization (Padlan 2008) requires the full structure of the antigen, trimer structures of the hemagglutinins in this study had to be obtained. Trimer structures of hemagglutinin are available in the Protein Data Bank (PDB), but the structures currently available in the PDB are all cleaved into HA1 and HA2. The exception is PDB Entry 1HA0, 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. We can only assume that the structure of the cleavage-site region in 1HA0 is identical to 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. PDB Entry 1HA0, therefore, is at the center of our analysis.

Modeling of the H3N2 hemagglutinin:
A model for the (2020-2021) H3N2 hemagglutinin was built using SWISS-MODEL (Peitsch 1995) (https://swissmodel.expasy.org) based on the first monomer of the 1HA0 structure and the sequence of GISAID Entry EPI_ISL_412381 (A/Hong Kong/2671/2019). The resulting model was used to generate a trimer using the 1HA0 trimer as the template. All the de-Antigenization calculations on H3N2 used this trimer.

Modeling of the H1N1 hemagglutinin:
A model for the H1N1 hemagglutinin was built using the PDB Entry 3UBQ. However, the 3UBQ structure is cleaved into HA1 and HA2 and the cleavage-site region had to be modeled first to complete the structure.

The regions around the cleavage sites in 3UBQ and 1HA0 are:

```
3UBQ:CPKYVKSTKLRLATGLRNIPSIQSRGLFGAIAGFIEG
     GWTGMVDGWYG
1HA0: CPKYVKQNTLKLATGMRNVPEKQTQGLFGAIAGFIE
     NGWEGMIDGWYG
```

Using the 1HA0 region as the template, the corresponding region in 3UBQ was built using SWISS-MODEL. Further, the HA1 and HA2 regions of 3UBQ were aligned with the corresponding regions of 1HA0 using TM-ALIGN (https://zhanglab.ccmb.med.umich.edu/TM-align) (Zhang and Skolnick 2005). The modeled cleavage-site region of 3UBQ was then spliced onto the HA1 and HA2 fragments of 3UBQ that had been aligned with the corresponding fragments of 1HA0. The modeled monomer of 3UBQ was then used as the template for modeling the (2020-2021) (GISAID EPI_ISL_391021 (A/Guangdong-Maonan/SWL1536/2019)) H1N1 hemagglutinin. The resulting model was then used to generate a trimer based on the 1HA0 trimer structure. All the de-Antigenization calculations on H1N1 used this trimer.

Table 4: Computed antigenicities of the cleavage-site epitopes prior to de-Antigenization and at various values of solvent accessibility used in the computations

		Exposure			
	(Prior)	0.50	0.3333	0.25	0.20
H3:	-1.41	0.96	1.54	1.87	2.14
H1:	-1.53	0.80	1.67	1.79	1.83
B:	0.46	2.00	2.10	2.27	2.37

The values given are standard deviations above the mean for the whole structure.

Table 5: Sequences of the modeled trimer H3 HA1 region prior to and after de-antigenization at various degrees of solvent exposure

STATLCLGHHAVPNGTIVKTITNDRIEVTNATELVQNSSIGEICDSPHQILDGGNCTLID	model trimer
STATLCLGHHAVPNGTIVTTITSSAIEVTTATELVQTSSIGAICGSPHTILDGGTCTLID	0.50
STATLCLGHHAVPNGTIVTTITSSAIEVTTATELVQTSSIGAICGSPHTILGGGTCTLID	0.3333
STATLCLGHHAVPNGTIVTTITSSAIEVTTATTLVQTSSIGAICGSPHTILGGGTCTLID	0.25
STATLCLGHHAVPNGTIVTTITSSAIEVTTATTLVQTSSIGAICGSPHTILGGGTCTLID	0.20

ALLGDPQCDGFQNKKWDLFVERSRAYSNCYPYDVPDYASLRSLVASSGTLEFKNESFNWA	model trimer
ALLGDPQCDGFQAKAWDLFVERSTAASGCYPYGVPTTASLRSLVASSGTLEFTSESFTWA	0.50
ALLGDPACSGFTAAAWDLFVERSTAASGCYPYGVPTTASLRSLVASSGTLEFTSESFTWA	0.3333
ALLGDPACSGATAAAWGLFVERSTAASGCYPYGVPTTASLRSLVASSGTLAFTSESFTWA	0.25
ALLGDPACSGATAAAWGLFVTTSTAASGCYPAGVPTTASLRSLVASSGTLAFTSESFTWA	0.20
GVTQNGKSFSCIRGSSSSFFSRLNWLTHLNYIYPALNVTMPNKEQFDKLYIWGVHHPVTD	model trimer
GVTQGGTSASCIRGSSSSFFSRLNWLTHLSAIYPALSVTMPNTETFDKLYIWGVHHPVTT	0.50
GVTTGGTSASCIRGSSSSFFSRLNWLTTLSAIYPALSVTMPNTTTASKLYIWGVHHPVTT	0.3333
GVTTGGTSASCITGSSSSFFSRLNWLTTLSAIYPALSVTMPNTTTASKLYIWGVHHPVTT	0.25
GVTTGGTSASCITGSSSSFFSRLNWLTTLSAIYPALSVTMPNTTTASKLYIWGVHHPVTT	0.20
KDQISLYAQSSGRITVSTKRSQQAVIPNIGSRPRIRNIPSRISIYWTIVKPGDILLINST	model trimer
TDQISLYAQSSGRITVSTTTSAQAVIPTIGSRPAIRSIPSRISIYWTIVTPGDILLITST	0.50
TGQISLYAASSGAITVSTTTSATAVIPTIGSRPAIRSIPSRISIYTTIVTPGDILLITST	0.3333
TGQISLYAASSGAITVSTTTSATAVIPTIGSRPAIASIPSRISITTTIVTPGDILLITST	0.25
TGQISLYAASSGAITVSTTTSATAVIPTIGSAPAIASIPSAISITTTIVTPGTILLITST	0.20
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPFQNVNRITYGACPRY	model trimer
GNLIAPRGYFKIRSGKSSIMASTAPIGTCTSECITPNGSIPNTKPFQNVNRITYGACPRY	0.50
GNLIAPTGYFTITSGTSSIMASTAPIGTCTSTCITPTGSIPTTTPFQGVTTITYGACPRT	0.3333
GNLIAPTGYFTITSGTSSIMASTAPIGTCTSTCITPTGSIPTTTPFQGVTTITYGACPTT	0.25
GNLIAPTGYFTITSGTSSIMASTAPIGTCTSTCITPTGSIPTTTPFQGVTTITYGACPTT	0.20
VKQSTLKLATGMRNVPEKQT	model trimer
VTQSTLKLATGMRNVPEKQT	0.50
VTQSTLTLATGMRNVPEKQT	0.3333
VTTSTLTLATGMRNVPEKQT	0.25
VTTSTLTLATGMRNVPEKQT	0.20

Shown are the sequences of the modeled H3 hemagglutinin (GISAID Entry EPI_ISL_412381 (A/Hong Kong/2671/2019)) HA1 region prior to and after de-antigenization at various degrees of solvent exposure. The number of sequence changes due to de-Antigenization at various degrees of exposure are 41 for 0.50 (half exposed), 71 for 0.3333 (one-third exposed), 80 for 0.25 (one-fourth exposed), and 87 for 0.20 (one-fifth exposed).

Table 6: Sequences of the modeled trimer H1 HA1 region prior to and after de-antigenization at various degrees of solvent exposure

DTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLRGVAPLHLGKCNIAGW	model trimer
STLCIGYHANNSTDTVTTVLTKNVTVTHSVTLLEDTHNGKLCKLRGVAPLHLGTCNIAGW	0.50
STLCIGYHANNSTDTVTTVLTTNVTVTTSVTLLETTTTGKLCTLTGVAPLTLGTCTIAGW	0.3333
STLCIGYHANNSTDTVTTVLTTNVTVTTSVTLLTTTTTGKLCTLTGVAPLTLGTCTIAGW	0.25
STLCIGYHANNSTDTVTTVLTTNVTVTTSVTLLTTTTTGTLCTLTGVAPLTLGTCTIAGW	0.20
ILGNPECESLSTARSWSYIVETSNSDNGTCYPGDFINYEELREQLSSVSSFERFEIFPKT	model trimer
ILGNPECASLSTARSWSYIVETSTSTNGTCYPGTFINYETLREQLSSVSSFERFEIFPKT	0.50
ILGNPACASLSTATSWSYIVETSTSTNGTCYPGTFITYTTLRAQLSSVSSFTRFEIFPTT	0.3333
ILGNPACASLSTATSWSYIVETSTSTTGTCYPGTFITTTTLRAALSSVSSFTRFTIFPTT	0.25
ILGAPACASLSTATSWSYIVETSTSTTGTCYPGTFITTTTLRAALSSVSSFTRTTIFPTT	0.20
SSWPNHDSDKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLNQTYINDKGKEVLVLWGIH	model trimer
SSWPTHDSTKGVTAACPTAGATSFYKNLIWLVKTGTSYPKLTQTYINDKGKEVLVLWGIH	0.50
SSWPTHDSTTGVTAACPTAGATSFYTNLIWLVKTGTSYPTLTATYINTTGTTVLVLWGIH	0.3333
SSWPTHDSTTGVTAACPTAGATSFYTNLIWLVTTGTSYPTLTATYINTTGTTVLVLWGIH	0.25
SSWPTHTSTTGVTAACPTAGATSFYTNLIWLVTTGTSYPTLTATYINTTGTTVLVLWGIH	0.20
HPPTIAVQESLYQNADAYVFVGTSRYSKKFKPEIATRPKVRDQEGRMNYYWTLVEPGDKI	model trimer
HPPTIAVQESLYQAATAYVTVGTSTYSKKFKPTIATRPTVRSQEGRMNYYWTLVTPGDKI	0.50
HPPTIAVQASLYAAATAYVTVGTSTYSKTFTPTIATRPTVASQTGRMTYYWTLVTPGDTI	0.3333
HPPTIAVQASLYAAATATVTVGTSTTSTTFTPTIATRPTVASTTGRMTYYWTLVTPGDTI	0.25
HPPTIAVQASLYAAATATVTVGTSTTSTTFTPTIATRPTVASTTGRMTYYWTLVTPGTTI	0.20

```
TFEATGNLVVPRYAFTMERDAGSGIIISDTPVHDCNTTCQTPEGAINTSLPFQNVHPITI      model trimer
TFEATGNLVVPRYAFTMERDAGSGIIISDTPVHTCTTTCQTPEGAITTSLPFQNVTPITI      0.50
TFEATGNLVVPRYAFTMTRTAGSGIIISTTPVTTCTTTCQTPEGAITTSLPFQNVTPITI      0.3333
TFEATGNLVVPRYAFTMTTTAGSGIIISTTPVTTCTTTCQTPTGAITTSLPFQNVTPITI      0.25
TFEATGNLVVPTYAFTMTTTAGSGIIISTTPVTTCTTTCQTPTGAITTSLPFQNVTPITI      0.20


GKCPKYVKSTKLRLATGLRNVPSIQS                                         model trimer
GKCPKYVKSTKLRLATGLRNVPSIQS                                         0.50
GTCPKYVTSTTLRLATGLRNVPSIQS                                         0.3333
GTCPKYVTSTTLTLATGLRNVPSIQS                                         0.25
GTCPKYVTSTTLTLATGLRNVPSIQS                                         0.20
```

Shown are the sequences of the modeled H1 hemagglutinin (GISAID Entry EPI_ISL_391021 (A/Guangdong-Maonan/SWL1536/2019)) HA1 region prior to and after de-antigenization at various degrees of solvent exposure. The number of sequence changes due to de-Antigenization at various degrees of exposure are 29 for 0.50 (half exposed), 68 for 0.3333 (one-third exposed), 82 for 0.25 (one-fourth exposed), and 87 for 0.20 (one-fifth exposed).

Table 7: Sequences of the modeled trimer B HA1 region prior to and after de-antigenization at various degrees of solvent exposure

```
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSHFANLKGTETRGKLCPKCLNC      model trimer
DRICTGITSSNSPHVVKTATQGEVGVTGVIPLTTTPTTSTFASLTGTETRGKLCPTCLGC      0.50

SAICTGITSSNSPTVVKTATQGTVGVTGVIPLTTTPTTSTFASLTGTTTRGTLCPTCLGC      0.3333
SAICTGITSSNSPTVVKTATQGTVGVTGVIPLTTTPTTSTFASLTGTTTRGTLCPTCLGC      0.25
SAICTGITSSNSPTVVKTATAGTVGVTGVIPLTTTPTTSTFASLTGTTTRGTLCPTCLGC      0.20

TDLDVALGRPKCTGKIPSARVSILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYEHVRLS      model trimer
TDLDVALGRPTCTGTIPSARVSILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYETVRLS      0.50

TDLDVALGRPTCTGTIPSAAVSILHEVTPVTSGCFPIMHSRTAIRQLPNLLRGYATVALS      0.3333
TDLDVALGAPTCTGTIPSAAVSILHEVTPVTSGCFPIMHSRTAIRQLPNLLRGAATVALS      0.25
TDLDVALGAPTCTGTIPSAAVSILHTVTPVTSGCFPIMHSRTAIRQLPNLLRGAATVALS      0.20

THNVINAEGAPGGPYKIGTSGSCPNITNGNGFFATMAWAVPDKNKTATNPLTIEVPYVCT      model trimer
TTNVIGATGAPGGPYTIGTSGSCPGITSGGGFFATMAWAVPTTTKTATSPLTIAVPYVCT      0.50

TTTVIGATGAPGGPYTIGTSGSCPGITSGGGFFATMAWAVPTTTKTATSPLTIAVPYVCT      0.3333
TTTVIGATGAPGGPYTIGTSGSCPGITSGGGFFATMAWAVPTTTTTATSPLTIAVPYVCT      0.25
TTTVIGATGAPGGPYTIGTSGSCPGITSGGGFFATMAVAVPTTTTTATSPLTIAVPYVCT      0.20

EGEDQITVWGFHSDNETQMAKLYGDSKPQKFTSSANGVTTHYVSQIGGFPNQTEDGGLPQ      model trimer
AGEDQITVWGFHSDTTTQMAALYGDSAPQKFTSSANGVTTHYVSQIGGFPGQTEDGGLPQ      0.50
AGEDQITVWGFHSDTTTAMAALYGDSAPQKFTSSANGVTTHYVSQIGGFPGQTEDGGLPQ      0.3333
AGESQITVWGFHSDTTTAMAALYGASAPQKFTSSANGVTTHYVSQIGGFPGATEDGGLPA      0.25
AGESQITVWGFHSDTTTAMAALYGASAPQTFTSSANGVTTHYVSQIGGFPGATADGGLPA      0.20

SGRIVVDYMVQKSGKTGTITYQRGILLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYGG      model trimer
SGRIVVDYMVQKSGKTGTITYQRGILLPQKVWCASGRSTVITGSLPLIGAADCLHEKYGG      0.50
SGRIVVDYMVQKSGKTGTITYQRGILLPQTVWCASGTSTVITGSLPLIGAADCLHETYGG      0.3333
SGRIVVDYMVQKSGKTGTITYTRGILLPTTVWCASGTSTVITGSLPLIGAASCLHETYGG      0.25
SGRIVVDYMVQTSGKTGTITYTRGILLPTTVWCASGTSTVITGSLPLIGAASCLHETYGG      0.20

LNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKE                      model trimer
LGTSKPYYTGEHAKAIGSCPIWVTTPLTLANGTKYRPPAKLLKE                      0.50

LGTSTPYYTGAHAKAIGSCPIWVTTPLTLANGTTYRPPAKLLKE                      0.3333
LGTSTPYYTGAHAKAIGSCPIWVTTPLTLANGTTYRPPAKLLKE                      0.25
LGTSTPAYTGAHAKAIGSCPIWVTTPLTLANGTTYRPPAKLLKE                      0.20
```

Shown are the sequences of the modeled B hemagglutinin (GISAID Entry EPI_ISL_257735 (B/Colorado/06/2017)) HA1 region prior to and after de-antigenization at various degrees of solvent exposure. The number of sequence changes due to de-Antigenization at various degrees of exposure are 35 for 0.50 (half exposed), 57 for 0.3333 (one-third exposed), 67 for 0.25 (one-fourth exposed), and 73 for 0.20 (one-fifth exposed).

Modeling of the B hemaglutinin:
A model for the B hemagglutinin was built using the PDB Entry 4M44. However, the 4M44 structure is also cleaved into HA1 and HA2 and the cleavage-site region had to be modeled to complete the structure. Here also the modeling made use of the 1HA0 structure.

The regions around the cleavage sites in 4M44 and 1HA0 are:

```
4M44: CPIWVKTPLKLANGTKYRPPAKL--
      LFFGAIAGFLEGGWEGMIAGW

1HA0:CPKYVKQNTLKLATGMRNVPEKQTQGLFGAIAGFIEN
      GWEGMIDGW
```

The two-residue difference between 4M44 and 1HA0 in these regions does not give us as much confidence in our modeling as in the case of H1N1. The results of the subsequent calculations, however, suggest that the results of the modeling of the cleavage-side region of 4M44 are plausible.

As with H1N1, the HA1 and HA2 fragments of 4M44 were first aligned with the corresponding fragments of 1HA0, then the modeled cleavage-site region was spliced onto these aligned fragments to generate a modeled 4M44 monomer, and the monomer used as template for the generation of a model for the (2018-2020) (GISAID EPI_ISL_257735 (B/Colorado/06/2017)) B hemagglutinin. (The presence of an unidentified residue, "X", in the QCG86174 (B/Washington/02/2019) sequence (see, Table 3) prevented the use of this (2020-2021) entry in the analysis. This is of no consequence since the residue at that position is mutated in the proposed B vaccine (see Table 10).). The resulting model of B hemagglutinin monomer was then used to generate a trimer based on the 1HA0 trimer structure. All the de-Antigenization calculations on B hemagglutinin used this trimer.

Calculation of antigenicities:
The method developed earlier (Padlan 2008) was used for the characterization of a putative antibody epitope centered at the cleavage site of the hemagglutinins. Briefly, the method identifies all the residues within a certain radius of the Cα (alpha-carbon) position of the cleavage site and assigns an antigenicity value for that site based on the summed reactivities of those residues weighted by their exposure to solvent. The reactivities that we used were those based on the values compiled by De Genst et al. (2002). The exposures to solvent were calculated by the method of Connolly (1983) using programs developed by Sheriff et al. (1985). There is a great variety in size and shape of antibody epitopes so that the exact extent of the cleavage-site epitope can only be guessed. Here, we will simply assume that the putative cleavage-site epitope comprises all the residues that are within 17 Angstroms (17A) of the alpha carbon (Cα) of the cleavage-site residue.

Trial calculations on the degree of de-Antigenization were performed to try 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). Mutations to lower the reactivity at the individual positions followed the replacement rules developed by one of us (Padlan 2008). The antigenicity computed for an epitope is expressed in terms of standard deviations above the mean for all the putative epitopes of the structure.

RESULTS AND DISCUSSION

The values of the antigenicity at various levels of solvent exposure for the putative cleavage-site epitope are shown in Table 4 for the H3, H1, and B hemagglutinins. It is remarkable that the antigenicity values prior to any de-Antigenization for the putative cleavage-site epitope for all three hemagglutiinin types are very low (Table 4). This is an obvious way that the influenza virus shields the cleavage site from antibody recognition during the immune response of the host. The antigenicities of the cleavage-site region are significantly increased by de-Antigenization of all three hemagglutinin types. This increase implies that the probability that the antibody response focusing on the cleavage-site region will also be increased. This can be expected to result in the interference with cleavage which is critical to viral fusion with the target-cell membrane. The de-Antigenized sequence that we propose could be used in the development of a potentially useful subunit vaccine, or as the basis for a transformed-viral vaccine, against influenza.

The sequences that we have chosen to propose to be useful in the development of possible subunit vaccines are the results of de-Antigenization at the level of one-fifth (0.20) exposed. (Those who wish to develop vaccines against flu might choose one of the other levels of de-Antigenization presented here.) Since it is the head region (HA1) that initially binds to the target cell, all the changes in this fragment are incorporated in the proposed vaccines. The conservation of almost all of HA2 is most probably for a (structural) reason. Thus, the unchanged portions of the HA2 fragment are also included in the proposed vaccines. Actually, a vaccine has been developed against influenza using just the HA2 fragment (Steel et al. 2010).

The results of our analyses are summarized in Tables 8, 9, and 10. For completeness, the antigenic sites of hemagglutinin proposed by Smith et al. (2004) for H3, by Caton et al. (1982) for H1, and by Stray and Pittman (2012) for B hemagglutinin are included in Tables 8, 9 and 10. The receptor-binding sites of H3 (Sauter et al. 1992), of H1 (Gamblin et al. 2004), and of B hemagglutinin (Ha et al. 2001) for B hemagglutinin are also indicated in the tables. The H3, H1, and B sequences that we propose as possible universal vaccines against seasonal influenza are marked "mut" in Tables 8, 9, and 10. A comparison of the "mut" and original sequences shows that the antigenic sites, but not the receptor-binding sites, are greatly affected by de-Antigenization.

Starting in the 2013-2014 (Northern Hemisphere) influenza season, WHO recommended the addition of another B influenza virus to form a quadrivalent vaccine. For the 2013-2014 and 2014-2015 seasons, WHO recommended a B/Massachusetts/2/2012-like virus (included in the list above) and a B/Brisbane/60/2008-like virus (also included above); for the 2015-2016, 2016-2017 and 2017-2018 seasons, WHO recommended a B/Brisbane/60/2008-like virus and a B/Phuket/3073/2013-like virus (also included in the list above); for the 2018-2019 and 2019-2020 seasons, WHO recommended a B/Colorado/06/2017-like virus (B/Victoria/2/87 lineage) (also in the above list) and a B/Phuket/3073/2013-like virus (B/Yamagata/16/88 lineage); for the 2020-2021 season, WHO recommended a B/Washington/02/2019 (B/Victoria lineage)-like virus and a B/Phuket/3073/2013 (B/Yamagata lineage)-like virus.

There are two types of B/Phuket/3073/2013-like viruses and these differ only at position 196 (Table 3 numbering); one type has an aspartic acid, while the other has an asparagine at this position. The residue at this position is mutated during the development of the universal B vaccine (see Table 10), so that this difference should not have any effect on the final proposed B vaccine. Similarly, there are two types of

B/Colorado/06/2017-like (B/Victoria/2/87 lineage) viruses and these differ also only at position 196; one type has a threonine, while the other has an asparagine. Again, since the residue at this position is changed during de-Antigenization, this difference should not make a difference in our B vaccine proposal. Thus, the vaccines that we propose make no distinction between trivalent and quadrivalent seasonal vaccines.

Table 8: Hemagglutinin H3 prior to and after de-Antigenization at 0.20 solvent exposure

```
        10        20        30        40        50        60
         |         |         |         | c  cc   |   e     |
STATLCLGHHAVPNGTIVKTITNDRIEVTNATELVQNSSIGEICDSPHQILDGGNCTLID
STATLCLGHHAVPNGTIVTTITSSAIEVTTATTLVQTSSIGAICGSPHTILGGGTCTLID      mut
               *       *           *  * *      *    *
        70        80        90        100       110       120
      e  |   ee    |         +   d     |         |   a a   |
ALLGDPQCDGFQNKKWDLFVERSRAYSNCYPYDVPDYASLRSLVASSGTLEFKNESFNWA
ALLGDPACSGATAAAWGLFVTTSTAASGCYPAGVPTTASLRSLVASSGTLAFTSESFTWA      mut
      *       *       **                              **     *
        130       140       150       160       170       180
  a a   a|    aaaa |    + bb b b   b   |   d d   |    +    b
GVTQNGKSFSCIRGSSSSFFSRLNWLTHLNYIYPALNVTMPNKEQFDKLYIWGVHHPVTD
GVTTGGTSASCITGSSSSFFSRLNWLTTLSAIYPALSVTMPNTTTASKLYIWGVHHPVTT      mut
  *   * ** * * **          ** ***          ***            * *
        190       200       210       220       230       240
bb  b+ bb|  d     d|    d   d|       + + d       |     d   |
KDQISLYAQSSGRITVSTKRSQQAVIPNIGSRPRIRNIPSRISIYWTIVKPGDILLINST
TGQISLYAASSGAITVSTTTSQTAVIPTIGSRPAIASIPSRISITTTIVTPGTILLITST      mut
*  *** * *   *         * *    *  ** * *     *
        250       260       270       280       290       300
         | e e     |       c c         |         |         |
GNLIAPRGYFKIRSGKSSIMRSDAPIGKCKSECITPNGSIPNDKPFQNVNRITYGACPRY
GNLIAPTGYFTITSGTSSIMASTAPIGTCTSTCITPTGSIPTTTPFQGVTTITYGACPTT      mut
                           *
        310       320       330       340       350       360
         |         |         |         |         |         |
VKQSTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLK
VTTSTLTLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLK      mut
  **            *                   *           *
        370       380       390       400       410       420
         |         |         |         |         |         |
STQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRVQDLEKYVEDTKIDLWSYNAELL
STQAAIAQINGKLNRLIGKTNEKFHQIEKEFSEVEGRVQDLEKYVEDTKIDLWSYNAELL      mut
      *         *                   *
        430       440       450       460       470       480
         |         |         |         |         |         |
VALENQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNETYDH
VALENQHTIDLTDSEMNKLFEATAKQLRENAEDMGNGCFKIYHKCDNACIGSIRNATYDH      mut
                     * *                          **    *
        490
         |
NVYRDEALNNRFQI
SVYRDEALNNRFQI                                                     mut
*
```

The sequence prior to de-Antigenizaton is that of the GISAID Entry EPI_ISL_412381 (A/Hong Kong/2671/2019). The sequence of the molecule that had been de-Antigenized at the 0.20 (one-fifth exposed) level (labeled "mut") is shown for comparison. The H3 antigenic sites, shown bold and underlined in the unmutated sequence, are from Smith et al. (2004). Residues in the antigenic sites are identified by the letters a, b, c, d, and e above them for the sites A, B, C, D, and E, respectively. Residues in the H3 receptor-binding site (Tyr90, Trp145, Thr147, His175, D182, Leu186, Thr218, and Ser220) are from Sauter et al. (1992) (PDB Entry 1HGH) and are indicated by plus signs (+) in red. When a receptor-binding site residue coincides with an antigenic-site residue, in particular residues at positions 147 and 182, the antigenic-site residue is shown in red.

Positions where any change had occurred from year to year in the period 1998-2021 are indicated by asterisks (*) below the sequences.

Table 9: Hemagglutinin H1 prior to and after de-Antigenization at 0.20 solvent exposure

```
        10        20        30        40        50        60
         |         |         |         |         |         |
DTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLRGVAPLHLGKCNIAGW
STLCIGYHANNSTDTVTTVLTTNVTVTTSVTLLTTTTTGTLCTLTGVAPLTLGTCTIAGW      mut
  *                              **      * * *   *  * **
        70        80        90        100       110       120
         ee eee    |         |+        |         |    e    |
ILGNPECESLSTARSWSYIVETSNSDNGTCYPGDFINYEELREQLSSVSSFERFEIFPKT
ILGAPACASLSTATSWSYIVETSTSTTGTCYPGTFITTTTLRAALSSVSSFTRTTIFPTT      mut
        * ****       *****       * **                      *
        130       140       150       160       170       180
   aa    |      d  d d       + +baab  aa aaa c   c        c+
SSWPNHDSDKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLNQTYINDKGKEVLVLWGIH
SSWPTHTSTTGVTAACPTAGATSFYTNLIWLVTTGTSYPTLTATYINTTGTTVLVLWGIH      mut
      ****  * * * * **   *  *  ** ***  * *** * * *        *
        190       200       210       220       230       240
 +   +  bb+ b b    |   c     |         |dd+      |      c  |
HPPTIAVQESLYQNADAYVFVGTSRYSKKFKPEIATRPKVRDQEGRMNYYWTLVEPGDKI
HPPTTAVQASLYAAATATVTVGTSTTSTTFTPTIATRPTVASTTGRMTYYWTLVTPGTTI      mut
  ***** *** ****   * ** *  ** *    *     **   *      *    *
        250       260       270       280       290       300
         |         |         |c        |         |         |
TFEATGNLVVPRYAFTMERDAGSGIIISDTPVHDCNTTCQTPEGAINTSLPFQNVHPITI
TFEATGNLVVPTYAFTMTTTAGSGIIISTTPVTTCTTTCQTPTGAITTSLPFQNVTPITI      mut
*   *   ** *   *** **     * ** *** ***   **    *      *  *
        310       320       330       340       350       360
         |         |         |         |         |         |
GKCPKYVKSTKLRLATGLRNVPSIQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSG
GTCPKYVTSTTLTLATGLRNVPSIQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSG      mut
 *     * *   **     *                        *
        370       380       390       400       410       420
         |         |         |         |         |         |
YAADLKSTQNAIDKITNKVNSVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWT
YAADLKSTQNAIAAITNKVNSVIEKMNTQFTAVGKEFNHLETRIENLNKKVDDGFLDIWT      mut
    *       **                          *  * *          *
        430       440       450       460       470       480
         |         |         |         |         |         |
YNAELLVLLENERTLDYHDSNVKNLYEKVRNQLKNNAKEIGNGCFEFYHKCDNTCMESVK
YNAELLVLLENERTLDAHDSNVKNLYEKVAAQLKNNAKEIGNGCFEFYHKCGGTCMESVK      mut
               *           **                    ***
        490
         |
NGTYDYPKYSEEAKLNR
NGTYDYPKYSEEAKLNR                                                  mut
           *
```

The sequence prior to de-Antigenizaton is that of the GISAID Entry EPI_ISL_391021 (A/Guangdong-Maonan/SWL1536/2019). The sequence of the molecule that had been de-Antigenized at the 0.20 (one-fifth exposed) level (labeled "mut") is shown for comparison. The H1 antigenic sites, shown bold and underlined, are from Caton et al. (1982). Residues in the antigenic sites are identified by the letters a, b, c, d, and e above them for sites Sa, Sb, Ca1, Ca2, and Ca3 of Caton et al. (1982), respectively. Residues in the H1 receptor-binding site (Tyr91, Trp150, Val152, His180, Pro183, Trp187, Leu198, and Gln223) are from Gamblin et al. (2004) (PDB Entry 1PVZ) and are indicated by plus signs (+) in red. Positions where any change had occurred from year to year in the period 1998-2021 are indicated by asterisks (*) below the sequences.

Table 10: Hemagglutinin B prior to and after de-Antigenization at 0.20 solvent exposure

```
        10        20        30        40        50        60
         |         |         |         |      cc |     e e |
DRICTGITSSNSPHVVKTATQGEVNVTGVIPLTTTPTKSHFANLKGTETRGKLCPKCLNC
SAICTGITSSNSPTVVKTATAGTVGVTGVIPLTTTPTTSTFASLTGTTTRGTLCPTCLGC      mut
                            *         *       *       *
        70        80        90        100       110       120
       ee|e e ee   cc        |    +    |         |     c   |
TDLDVALGRPKCTGKIPSARVSILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYEHVRLS
TDLDVALGAPTCTGTIPSAAVSILHTVTPVTSGCFPIMHSRTAIRQLPNLLRGAATVALS      mut
          * * **   **      *                   *       **
        130       140       150       160       170       180
```

```
dd  dddd |     aa  |a    aaaaa   a   + |  bbbbbb |         |
THNVINAEGAPGGPYKIGTSGSCPNITNGNGFFATMAWAVP-DKNKTATNPLTIEVPYVC
TTTVIGATGAPGGPYTIGTSGSCPGITSGGGFFATMAVAVP-TTTTTATSPLTIAVPYVC      mut
**   *  *   *  **        * ***           **** *   *  *   **
        190       200       210       220       230       240
 ddd     |  +   bbbb bb+   bb|         |         |         |
TEGEDQITVWGFHSDNETQMAKLYGDSKPQKFTSSANGVTTHYVSQIGGFPNQTEDGGLP
TAGESQITVWGFHSDTTTAMAALYGASAPQTFTSSANGVTTHYVSQIGGFPGATADGGLP      mut
***             **** **     *          *         *  *
        250       260       270       280       290       300
         |dd  dddd |         |        c|   c     |         |
QSGRIVVDYMVQKSGKTGTITYQRGILLPQKVWCASGRSKVIKGSLPLIGEADCLHEKYG
ASGRIVVDYMVQTSGKTGTITYTRGILLPTTVWCASGTSTVITGSLPLIGAASCLHETYG      mut
          *  **     *    *                                *
        310       320       330       340       350       360
         |         |         |         |         |         |
GLNKSKPYYTGEHAKAIGNCPIWVKTPLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGW
GLGTSTPAYTGAHAKAIGSCPIWVTTPLTLANGTTYRPPAKLLKERGFFGAIAGFLEGGW      mut
           *
        370       380       390       400       410       420
         |         |         |         |         |         |
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH
EGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELH      mut

        430       440       450       460       470       480
         |         |         |         |         |         |
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVEIG
NEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERKLKKMLGPSAVEIG      mut
                                                        *
        490       500       510
         |         |         |
NGCFETKHKCNQTCLDKIAAGTFDAGEFSLPTFDSLNI
NGCFETKHKCNQTCLDAIAAGTFAAGEFSLPTFDSLNI                            mut
               *       *
```

The sequence prior to de-Antigenizaton is that of the GISAID Entry EPI_ISL_257735 (B/Colorado/06/2017). The sequence of the molecule that had been de-Antigenized at the 0.20 (one-fifth exposed) level (labeled "mut") is shown for comparison. The B antigenic sites, shown bold and underlined, are from Stray and Pittman (2012). Residues in the antigenic sites are identified by the letters a, b, c, d, and e above them for sites BA, BB1+BB2, BC, BD, and BE of Stray and Pittman (2012), respectively. The location of the antigenic residues beyond position 170 differs by one from those reported by Stray and Pittman (2012) because of the variation in the insertions at positions 161-165 (see Table 3). Residues in the B receptor-binding site (Phe95, Trp158, His191, and Tyr202) are from Wang et al. (2007) (PDB Entry 2RFU) and are indicated by plus signs (+) in red. Positions where any change had occurred from year to year in the period 1998-2021 are indicated by asterisks (*) below the sequences.

CONFLICT OF INTEREST

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.

CONTRIBUTION OF INDIVIDUAL AUTHORS

C. P. Mikita provided the medical significance of the work and researched the early literature. E. A. Padlan did the sequence 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; 16:548-58.

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; 277:29897-907.

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.

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.

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; 82:1104-07.

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):371-6

Steel J, Lowen AC, Wang TT, Yondola M, Gao Q, Haye K, Garcia-Sastre A, Palese P. Influenza virus vaccine based on the conserved hemagglutinin stalk domain. mBio 2010; 1(1):e00018-10.

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.

Wang Q, Tian X, Chen X, Ma J. Structural basis for receptor specificity of influenza B virus hemagglutinin. Proc Natl Acad Sci USA 2007; 104(43):16874–16879. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Re

Wiley DC, Wilson IA, Skehel JJ. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature 1981; 289:373-8.

Wilson IA, Skehel JJ, Wiley DC. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 1981; 289:366–73.

Zhang Y, Skolnick J. TM-align: A protein structure alignment algorithm based on TM-score. Nucleic Acids Res 2005; 33:2302-9.