

## Research Article

# ***In Silico* Identification of Highly Conserved Epitopes of Influenza A H1N1, H2N2, H3N2, and H5N1 with Diagnostic and Vaccination Potential**

**José Esteban Muñoz-Medina,<sup>1</sup> Carlos Javier Sánchez-Vallejo,<sup>2</sup> Alfonso Méndez-Tenorio,<sup>2</sup> Irma Eloísa Monroy-Muñoz,<sup>3</sup> Javier Angeles-Martínez,<sup>4</sup> Andrea Santos Coy-Arechavaleta,<sup>1</sup> Clara Esperanza Santacruz-Tinoco,<sup>1</sup> Joaquín González-Ibarra,<sup>5</sup> Yu-Mei Anguiano-Hernández,<sup>1</sup> César Raúl González-Bonilla,<sup>5</sup> Eva Ramón-Gallegos,<sup>6</sup> and José Alberto Díaz-Quiñonez<sup>7,8</sup>**

<sup>1</sup>Laboratorio Central de Epidemiología, Centro Médico Nacional La Raza, Instituto Mexicano del Seguro Social, Avenida Jacarandas S/N, Esquina Circuito Interior, Colonia La Raza Delegación Azcapotzalco, 02990 México, DF, Mexico

<sup>2</sup>Laboratorio de Biotecnología y Bioinformática Genómica, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Carpio y Plan de Ayala S/N, Colonia Plutarco Elías Calles, Delegación Miguel Hidalgo, 11340 México, DF, Mexico

<sup>3</sup>Laboratorio de Genómica, Departamento de Genética y Genómica Humana, Instituto Nacional de Perinatología "Isidro Espinosa de los Reyes", Montes Urales 800, Colonia Lomas Virreyes, Delegación Miguel Hidalgo, 11000 México, DF, Mexico

<sup>4</sup>Laboratorio de Genómica, Departamento de Biología Molecular, Instituto Nacional de Cardiología "Ignacio Chávez", Juan Badiano No. 1, Colonia Sección XVI, Delegación Tlalpan, 14080 México, DF, Mexico

<sup>5</sup>División de Laboratorios de Vigilancia e Investigación Epidemiológica, Instituto Mexicano del Seguro Social, Mier y Pesado No. 120, Colonia Del Valle, Delegación Benito Juárez, 03100 México, DF, Mexico

<sup>6</sup>Laboratorio de Citopatología Ambiental, Departamento de Morfología, Escuela Nacional de Ciencias Biológicas Campus Zacatenco, Instituto Politécnico Nacional, Wilfrido Massieu S/N, Colonia Nueva Industrial Vallejo, Delegación Gustavo A. Madero, 07738 México, DF, Mexico

<sup>7</sup>Instituto de Diagnóstico y Referencia Epidemiológicos "Manuel Martínez Báez", Francisco de P. Miranda No. 177, Colonia Unidad Lomas de Plateros, Delegación Álvaro Obregón, 01480 México, DF, Mexico

<sup>8</sup>Facultad de Medicina, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Colonia Copilco Universidad, Delegación Coyoacán, 04510 México, DF, Mexico

Correspondence should be addressed to José Alberto Díaz-Quiñonez; [alberto.diaz@salud.gob.mx](mailto:alberto.diaz@salud.gob.mx)

Received 22 February 2015; Revised 13 June 2015; Accepted 14 June 2015

Academic Editor: Swapnil Sinha

Copyright © 2015 José Esteban Muñoz-Medina et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The unpredictable, evolutionary nature of the influenza A virus (IAV) is the primary problem when generating a vaccine and when designing diagnostic strategies; thus, it is necessary to determine the constant regions in viral proteins. In this study, we completed an *in silico* analysis of the reported epitopes of the 4 IAV proteins that are antigenically most significant (HA, NA, NP, and M2) in the 3 strains with the greatest world circulation in the last century (H1N1, H2N2, and H3N2) and in one of the main aviary subtypes responsible for zoonosis (H5N1). For this purpose, the HMMER program was used to align 3,016 epitopes reported in the Immune Epitope Database and Analysis Resource (IEDB) and distributed in 34,294 stored sequences in the Pfam database. Eighteen epitopes were identified: 8 in HA, 5 in NA, 3 in NP, and 2 in M2. These epitopes have remained constant since they were first identified (~91 years) and are present in strains that have circulated on 5 continents. These sites could be targets for vaccination design strategies based on epitopes and/or as markers in the implementation of diagnostic techniques.

## 1. Introduction

The influenza A virus (IAV) appears seasonally, causes annual epidemics, and occasionally presents a new strain with pandemic reach, leading to severe consequences for global health and for the global economy [1, 2]. Every year, influenza affects approximately 15% of the world population, which translates to 3 to 5 million infections and 500,000 deaths [3, 4].

IAV is an enveloped virus and a member of the orthomyxoviridae family; its genome consists of eight segments of simple chain RNA of negative polarity that code for 3 structural proteins (HA, NA, and M2), 1 membrane protein (M1), and 6 nonstructural proteins (NS1, NEP/NS2, PA, PB1, PB1-F2, and PB2) [5].

IAV is classified according to its two principal membrane antigens: HA (18 subtypes) and NA (11 subtypes) [6, 7]. In humans, 6 subtypes (H1, H2, H3, H5, H7, and H9) [8–10] have been detected; however, only 3 of those have crossed the species barrier and have the capacity to transmit from human to human (H1N1, H2N2, and H3N2) [11].

The following three subtypes are responsible for the pandemics of the past century: H1N1 (Spanish flu, 1918), H2N2 (Asian flu, 1957), and H3N2 (Hong Kong flu, 1968) [12].

The last influenza pandemic occurred in 2009 due to a new pandemic virus A H1N1 (A H1N1pdm 09). This virus was detected in 214 cities on 5 continents and, up to July 30, 2010, had caused 18,389 cases [4].

These pandemic viruses arose due to a combination of aviary, pig, and human influenza viruses [13–15] because, in contrast to other respiratory viruses, they present two mutation mechanisms: genetic and antigenic drift [16].

These mutation mechanisms confer an unpredictable, evolving character to the influenza viruses, which is the principal difficulty to overcome when designing a vaccine. This difficulty occurs because while vaccination has been an effective method against 60 to 90% of seasonal strains [17], it has not been effective against pandemic viruses [18].

Inactivated trivalent vaccines that contain the hemagglutinin (HA) protein of the influenza A (H1N1) and A (H3N2) viruses and influenza B virus are the only authorized commercial vaccines [19]. Due to the high mutation rate of HA, it is necessary to vaccinate every year, according to World Health Organization suggestions [20].

Antiviral strategies that are currently in development include neutralizing antibodies [21], small molecule inhibitors [22], and universal vaccines [23] against the influenza virus based on conserved epitopes. Universal vaccines are used as an alternative approach for improving immunogenicity and cross-protection against emerging strains, shortening production time, and reducing side effects [24–26].

Therefore, it is of vital importance to know the sites of the proteins or antigenic determinants among the different strains of the influenza virus that historically have been recognized by the immune system. Therefore, the present study completed an *in silico* identification of highly conserved epitopes with diagnostic and vaccination potential in the HA, NA, MP, and M2 proteins of the influenza virus that have been

reported from 1918 to 2014 for the primary strains that have circulated in the world (H1N1, H2N2, H3N2, and H5N1).

## 2. Materials and Methods

**2.1. In Silico Search and Attainment of Epitopes.** To complete the epitope search on the IEDB site (<http://www.iedb.org/>), the following inclusion criteria were taken into account: experimentally reported epitopes for HA, NA, NP, and M2 proteins corresponding to the H1N1, H2N2, H3N2, and H5N1 subtypes of influenza A from 1918 up to 2011.

**Exclusion Criteria.** Epitopes corresponding to other influenza subtypes or proteins or that had not been verified experimentally were excluded.

**Elimination Criteria.** All sequences shorter than 7 amino acids or with erroneous information were eliminated; individual archives were created in FASTA format using the SeqBuilder program (DNASTAR Madison, Wisconsin, USA).

**2.2. Sequence Download.** Using the Pfam database (<http://pfam.xfam.org/>), sequences corresponding to the hemagglutinin (PF00509), neuraminidase (PF00064), nucleoprotein (PF00506), and matrix 2 (PF00599) protein families were downloaded from the Pfam-A entries along with sequences from the NCBI database, representing manually refined, high-quality families.

**2.3. Construction of Hidden Markov Models and Epitope Alignments.** Using the *hmmbuild* algorithm of the HMMER program (European Bioinformatics Institute), each one of the alignments downloaded from the Pfam database was used in the construction of hidden Markov models. This algorithm generates a file with the extension \*.hmm that contains the consensus sequence for each family of proteins.

Subsequently, the *hmmalign* algorithm was used to align the epitopes with their corresponding hidden Markov models (\*.hmm). This algorithm generates a file with the extension \*.sto that contains the epitopes aligned with the consensus sequence for each protein.

**2.4. Epitope Selection.** Based on the files generated by the *hmmalign* program, the regions that presented the greatest frequency of reports of epitopes, at least 2 standard deviations above the mean, were selected for each protein and strain analyzed. Subsequently, two new alignments were completed with the help of the *Megalign* program (DNASTAR Madison, Wisconsin, USA). The first alignment was between the present epitopes in each region to determine the consensus recognition sequence. The second was between this consensus sequence and some current strains that circulated between 2013 and 2014 to determine the similitude percentage with strains that currently circulate.

**2.5. Protein Modeling.** To complete the modeling of the consensus sequences generated using the *hmmbuild* program,

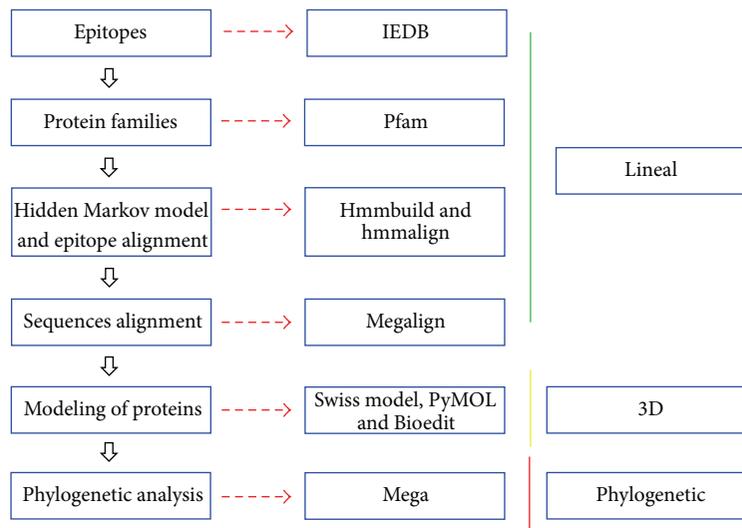


FIGURE 1: Flowchart of the stages and programs utilized in this study.

mold structures downloaded from the Protein Data Bank (PDB) with similitude percentages  $\geq 70\%$  were utilized. With these structures as a template, a three-dimensional structure of each protein was modeled in the Swiss-Model (Swiss Institute of Bioinformatics Biozentrum, University of Basel, Switzerland) virtual platform; the consensus regions of each epitope were highlighted in the model utilizing the PyMOL program (Schrödinger K. K., Japan).

**2.6. Phylogenetic Analysis.** For each epitope group, a phylogenetic tree with the complete sequence of the protein to which it belongs was constructed using the MEGA program [27]; the tree was constructed with the Neighbor-Joining algorithm and with a bootstrap of 1000 replicas.

**2.7. Strain Circulation Analysis.** To determine the reach that using these epitopes would have, in terms of vaccination or diagnosis in the world population, an analysis of the strains containing the consensus epitopes determined in this study was completed; for each epitope, data from viruses isolated from cases that were reported were collected.

The stages of the entire analysis and the programs used are summarized in Figure 1.

**2.8. Statistical Analysis.** Descriptive statistics were used to determine the regions of the proteins that were recognized with the greatest frequency. For this purpose, the protein was divided into groups, each containing 10 amino acids, the frequencies of epitopes present in each group were observed, and the means and standard deviations were calculated. Up to 3 regions with frequencies at least 2 standard deviations above the mean were chosen for each protein.

### 3. Results

**3.1. Distribution of Epitopes and Sequences.** In total, 3,016 epitopes and 34,294 sequences were identified. Of the

epitopes and sequences, respectively, 1,352 and 15,102 belonged to H1N1, 91 and 412 belonged to H2N2, 1,006 and 12,186 belonged to H3N2, and, finally, 567 and 6,594 belonged to H5N1. In all cases, HA is the protein for which more epitopes and sequences have been reported (Table 1).

**3.2. Identification of the Region with the Greatest Frequency of Antigenic Recognition.** Based on the frequency distribution graph generated using the *hmalign* program, 25 sites were identified which showed the greatest frequency to which the epitopes aligned. These sites were distributed in the following manner: 8 for H1N1, 3 for H2N2, 6 for H3n2, and 8 for H5N1 (Table 2). All of these highly recognized regions had frequencies at least 3 standard deviations above the mean. An example of the graph generated by the *hmalign* program is shown in Figure 2.

**3.3. Determination of Consensus Epitopes.** From the epitopes present in the 25 identified sites, 18 consensus epitopes generated using the Megalign program were obtained. These sites were distributed in the following manner: 6 for H1N1, 3 for H2N2, 3 for H3N2, and 6 for H5N1.

Upon completing the second alignment to determine the similitude percentages with strains that are currently circulating, it was observed that homology existed between 10 of the 18 consensus epitopes. In all cases, the similitude percentages were greater than 90% (Table 3 and Figure 3).

**3.4. Three-Dimensional Arrangement of the Consensus Epitopes.** With the objective of observing the level of exposure of the consensus epitopes in the three-dimensional protein structures, a mold was downloaded from PDB for each case with a similitude percentage  $\geq 70\%$  with the consensus sequence (Table 4).

In the case of the HA protein, 5 of the 8 consensus epitopes were located in the globular zone, very close to the sialic acid-binding site, while the remaining epitopes were located in the

TABLE 1: Distribution of epitopes and sequences obtained for each subtype.

		Viral strains			
		H1N1 <i>n</i> (%)	H2N2 <i>n</i> (%)	H3N2 <i>n</i> (%)	H5N1 <i>n</i> (%)
Proteins	HA				
	E	686 (50.7)	43 (47.3)	525 (52.2)	384 (67.8)
	S	8416 (55.7)	175 (42.5)	7313 (60.0)	3024 (45.9)
	NA				
	E	283 (20.9)	0 (0)	195 (19.4)	126 (22.2)
	S	4462 (29.6)	129 (31.3)	2909 (23.9)	1981 (30.0)
	NP				
	E	328 (24.3)	47 (51.6)	252 (25.0)	45 (7.9)
	S	1225 (8.1)	78 (18.9)	1096 (9.0)	1019 (15.5)
	M2				
E	55 (4.1)	1 (1.1)	34 (3.4)	12 (2.1)	
S	999 (6.6)	30 (7.3)	868 (7.1)	570 (8.6)	

Note: S = sequences and E = epitopes.

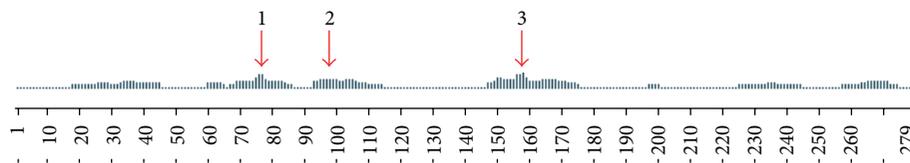


FIGURE 2: Example of the frequency distribution graph generated using the *hmmalign* program for hemagglutinin of the H2N2 subtype. The arrows indicate the selected regions for this protein.

TABLE 2: Sites with the greatest frequency of epitopes.

Subtype	Protein	Site
H1N1	HA	123–140; 205–219; 330–352
	NA	93–109; 204–222
	NP	414–427
	M2	25–43; 48–62
H2N2	HA	200–220; 251–274; 332–356
H3N2	HA	112–131; 322–336;
	NA	143–160; 311–331
	NP	52–69; 363–378
H5N1	HA	174–195; 227–246; 247–267
	NA	103–121; 231–252
	NP	190–211; 257–278; 400–418

stem region. For the NA protein, all of the defined epitopes were separate from the zanamivir-binding site. Finally, in the case of the M2 protein, consensus epitopes were identified in the membrane and transmembrane regions (Figure 4).

**3.5. Phylogenetic Analysis.** Phylogenetic analysis revealed the evolutionary distance between the sequence in which each epitope was identified for the first time and the sequence of either the 2013–2014 season vaccine strain (in the case of H1N1 and H3N2) or a strain reported in 2013 (in the case of H5N1). The mutation rates for nucleotide substitutions per site per

year were  $1.2 \times 10^{-3}$  for HA H1N1,  $3.5 \times 10^{-3}$  for HA H3N2,  $3.4 \times 10^{-3}$  for HA H5N1,  $1.6 \times 10^{-3}$  for NA H1N1,  $1.1 \times 10^{-4}$  for NA H5N1, and  $7.8 \times 10^{-4}$  for M2 H1N1. For the HA H2N2, NA H2N2, and NP H5N1 proteins, a phylogenetic tree was not completed because the isolate sequences were sourced from 3 or fewer strains (Figure 5).

**3.6. Geographic Distribution of Consensus Epitopes.** The circulation analysis revealed that the epitopes defined in this study have been present in isolates obtained from 5 continents: the H1N1 epitopes have been identified in China, Mongolia, the USA, France, and Puerto Rico; the H2N2 epitopes have only been identified in Japan; the H3N2 epitopes were isolated in Panama, Argentina, Canada, the USA, China, Holland, France, and Australia; and, finally, the H5N1 epitopes were observed in Indonesia, Thailand, Vietnam, China, Japan, France, and Egypt (Figure 6).

## 4. Discussion

While the interest in the scientific community is focused on the production of an influenza vaccine that provides cross-protection against multiple subtypes [23], thus far, the vaccine must still be modified each year [20].

This necessity emphasizes the importance of identifying IAV sequence regions that remain constant and that, moreover, have the capacity to induce immunological responses from T and B cells. Therefore, this study completed an *in*

TABLE 3: Consensus epitopes.

Subtype	Protein	Epitope	Sequence	T or B cell	Reference
H1N1	HA	1	<sup>127</sup> SVSSFERFEIFPK <sup>136</sup>	T	[28]
		2	<sup>333</sup> VTGLRNIPSIQSRGL <sup>346</sup>	T y B	[29, 30]
	NA	3	<sup>96</sup> GWAIYSKDNN <sup>106</sup>	T	[31]
		4	<sup>206</sup> LKYNGIITETIKSW <sup>219</sup>	T	[32]
	M2	5	<sup>29</sup> AASIIILHLIL <sup>40</sup>	T	[33]
		6	<sup>51</sup> ITRLFKHGLK <sup>60</sup>	T	[32]
H2N2	HA	7	<sup>204</sup> LYQNVGTYVSVGTST <sup>218</sup>	T	[34]
		8	<sup>255</sup> FESTGNLIAPEYGFKIS <sup>271</sup>	T	[28]
		9	<sup>337</sup> IESRGLFGAIAAGFIEGGWQ <sup>355</sup>	T	[35]
H3N2	HA	10	<sup>113</sup> CYPYDVPDYASLRS <sup>126</sup>	T y B	[32, 36]
		11	<sup>324</sup> YVKQNTLKLA <sup>333</sup>	T y B	[37, 38]
	NA	12	<sup>314</sup> SSYVCSGLVGDTPR <sup>327</sup>	T	[39]
H5N1	HA	13	<sup>229</sup> IATRSKVNQSGRM <sup>243</sup>	T y B	[32, 40]
		14	<sup>105</sup> SHLECRTFFLTQGALLNDKH <sup>124</sup>	T y B	[41, 42]
	NA	15	<sup>234</sup> KIFKMEKGVVKSVEL <sup>249</sup>	T y B	[32, 41]
		16	<sup>192</sup> ELIRMIKRGINDRNFWR <sup>208</sup>	T	[43]
	NP	17	<sup>261</sup> RSALILRGVAHKSC <sup>276</sup>	T	[39]
		18	<sup>401</sup> ASAGQISVQPTFSVQRN <sup>417</sup>	T	[39]

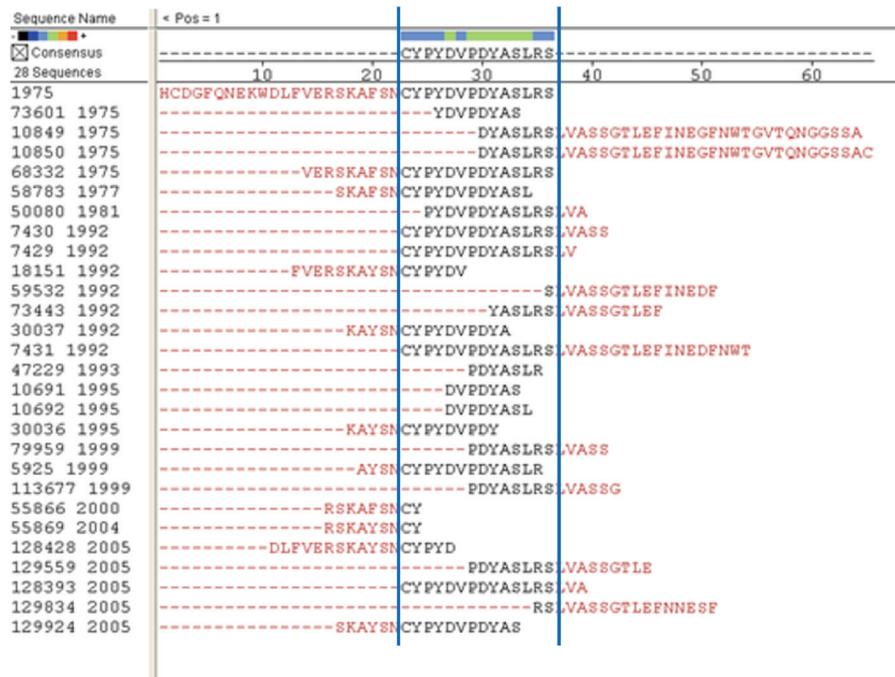


FIGURE 3: Example of the epitope alignment completed using the Megalign program. The depicted figure corresponds to the hemagglutinin of the H3N2 subtype. The blue lines show the consensus sequences in the different epitopes reported.

*in silico* analysis using previously reported epitopes and those for which evidence indicates they are capable of inducing an immunological response.

As was to be expected, given that H1N1 was the first IAV subtype that was recognized and has been circulating in the world since 1918, of the epitopes used in this study, those corresponding to H1N1 represented 44.8%, while those

corresponding to H2N2 only represented 3%. This difference could, in part, be due to the fact that H2N2 only circulated after the 1957 pandemic up until 1968, when it was displaced by the H3N2 subtype [44, 45].

However, for all of the subtypes, with the exception of H2N2, the protein in which the greatest number of epitopes has been described is HA, with the following percentages:

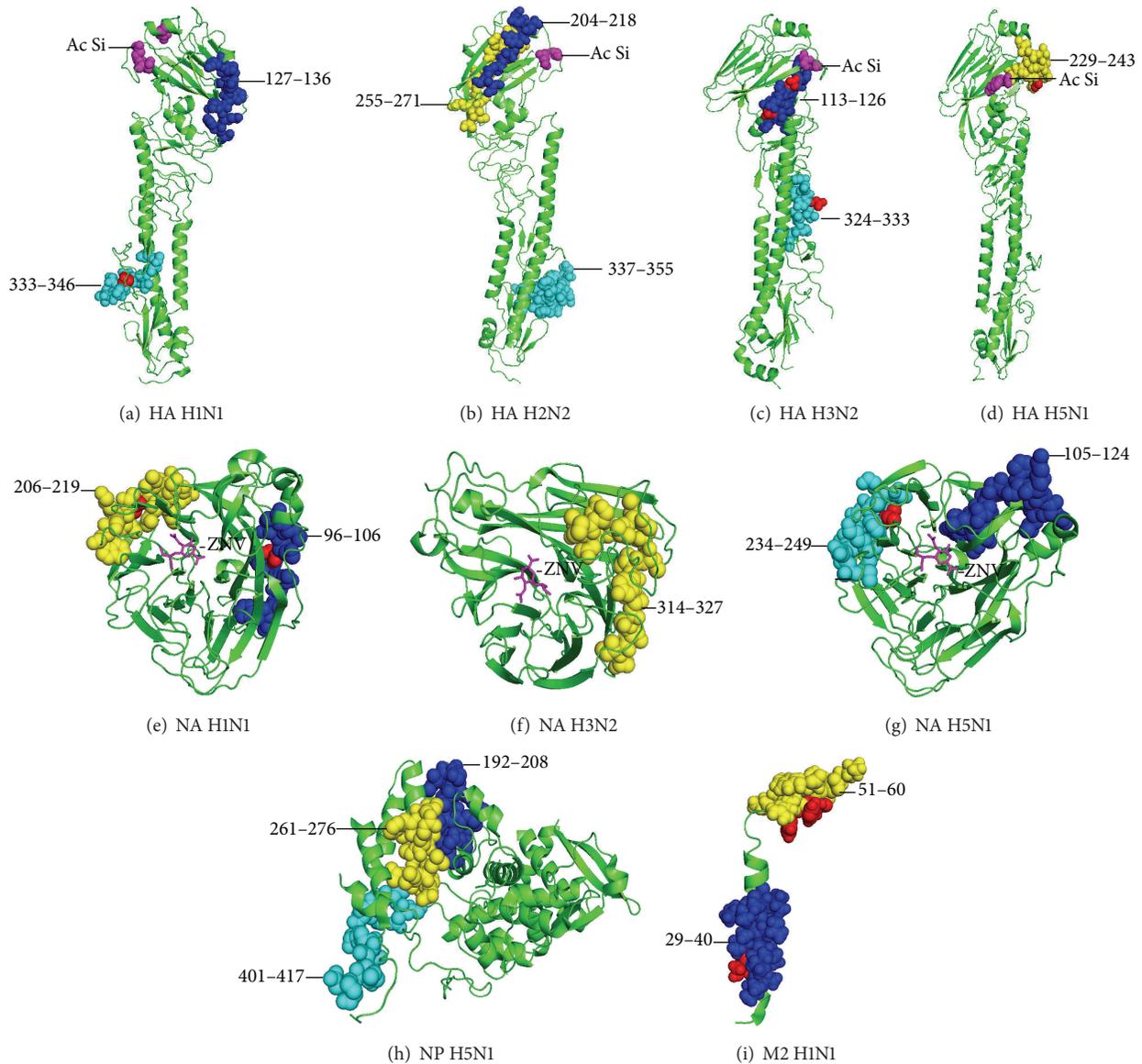
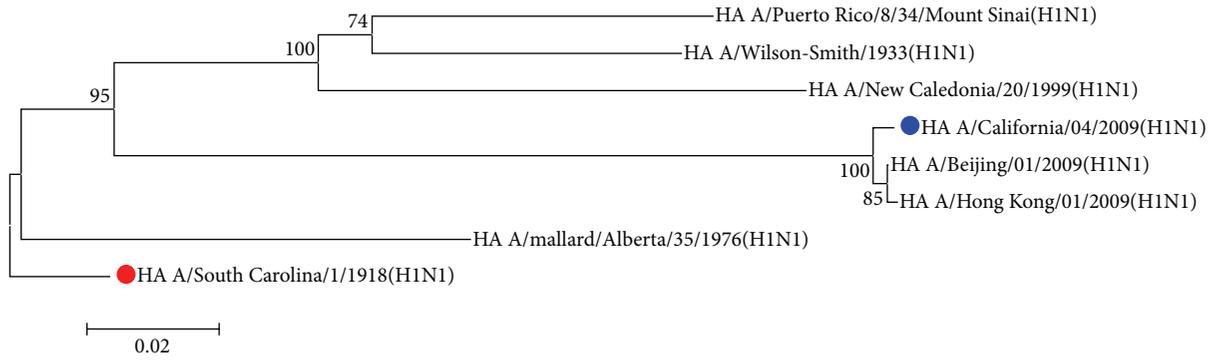


FIGURE 4: Three-dimensional localization of consensus epitopes. The protein structures are shown in green, and the conserved epitopes are shown as blue, yellow, and cyan spheres. The mutations within the consensus epitopes, compared to the current strains, are shown as red spheres. (a) [PDB: 2WR3] HA H1N1 in blue <sup>127</sup>SVSSFERFEIFPK<sup>136</sup>; in cyan <sup>333</sup>VTGLRNIPSIQSRGL<sup>346</sup> (b) [PDB: 2WRD] HA H2N2 in blue <sup>204</sup>LYQNVGTIVSVGTST<sup>218</sup>; in yellow <sup>255</sup>FESTGNLIAPEYGFKIS<sup>271</sup>; in cyan <sup>337</sup>IESRGLFGAIAFGIEGGWQ<sup>355</sup> (c) [PDB: 2YP2] HA H3N2 in blue <sup>113</sup>CYPYDVPDYASLRS<sup>126</sup>; in cyan <sup>324</sup>YVKQNTLKLKLA<sup>333</sup> (d) [PDB: 4BGW] HA H5N1 in yellow <sup>229</sup>IATRSKVNQSGRM<sup>243</sup> (e) [PDB: 4B7J] NA H1N1 in blue <sup>96</sup>GWAIYSKDNNS<sup>106</sup>; in yellow <sup>206</sup>LKYNHITETIKSW<sup>219</sup> (f) [PDB: 4GZQ] NA H3N2 in yellow <sup>314</sup>SSVCSGLVGDTPR<sup>327</sup> (g) [PDB: 4B7J] NA H5N1 in blue <sup>105</sup>SHLECRFTFLTQGALLNDKH<sup>124</sup>; in cyan <sup>234</sup>KIFKMEKGVVKSVEL<sup>249</sup> (h) NP H5N1 in blue <sup>192</sup>ELIRMIKRGINDRNFWR<sup>208</sup>; in yellow <sup>261</sup>RSALILRGSVAHKSL<sup>276</sup>; in cyan <sup>401</sup>ASAGQISVQPTFSVQRN<sup>417</sup> (i) [PDB: 2RLF] M2 H1N1 in blue <sup>29</sup>AASIIILHLIL<sup>40</sup>; in yellow <sup>51</sup>ITRLFKHGLK<sup>60</sup>.

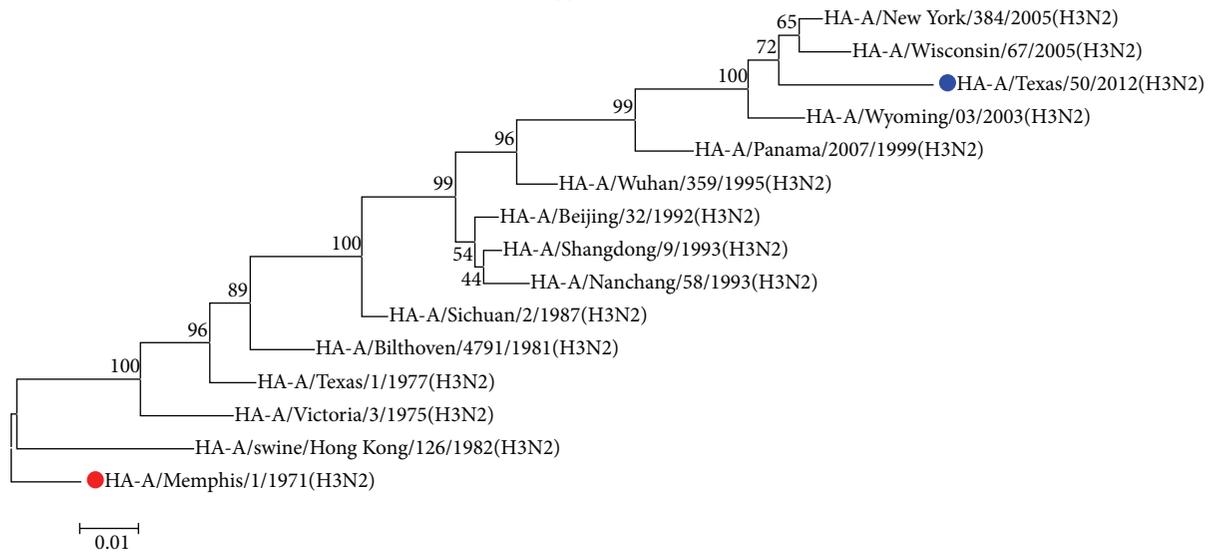
H1N1 50.7%, H3N2 52.2%, and H5N1 67.7%. This finding is similar to observations made by Bulimo and Cols in 2012 [46]. Interestingly, M2 is the protein for which fewer epitopes have been described, even though the M2 protein has been one of the main targets in studies focusing on vaccine development in the last few years [47–50].

It is important to note that although the H5N1 subtype has only been detected sporadically in humans, it is generally

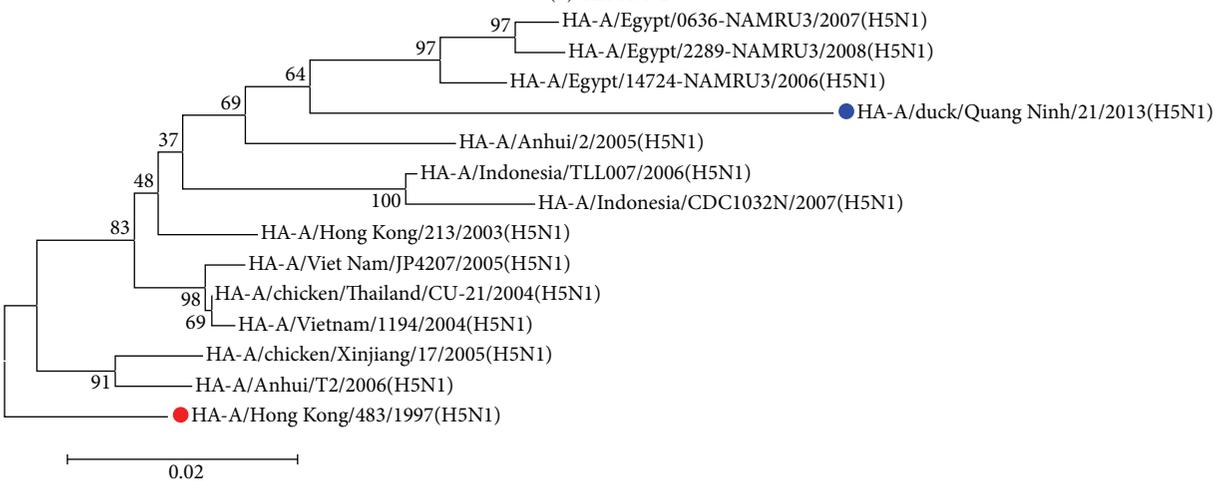
associated with cases of zoonosis in individuals who have very close contact with poultry or undomesticated birds as a result of work activities or hunting, respectively [51]. The epitopes reported for the H5N1 subtype represent 17.6% of the total, which tells us that it is worthy of study, as its high pathogenicity has been widely observed in both birds [52, 53] and humans [54, 55] since the first cases noted in 2003 [56]. Moreover, many authors believe that H5N1 could be the next



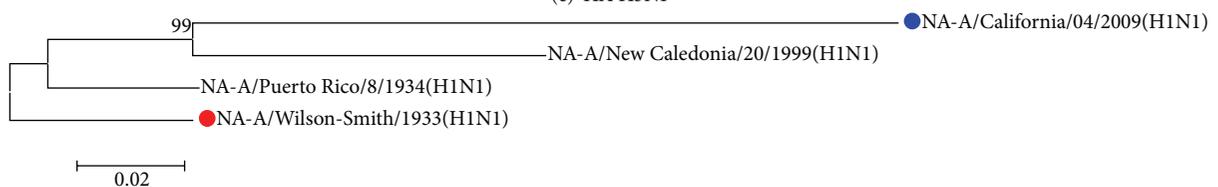
(a) HA H1N1



(b) HA H3N2



(c) HA H5N1



(d) NA H1N1

FIGURE 5: Continued.

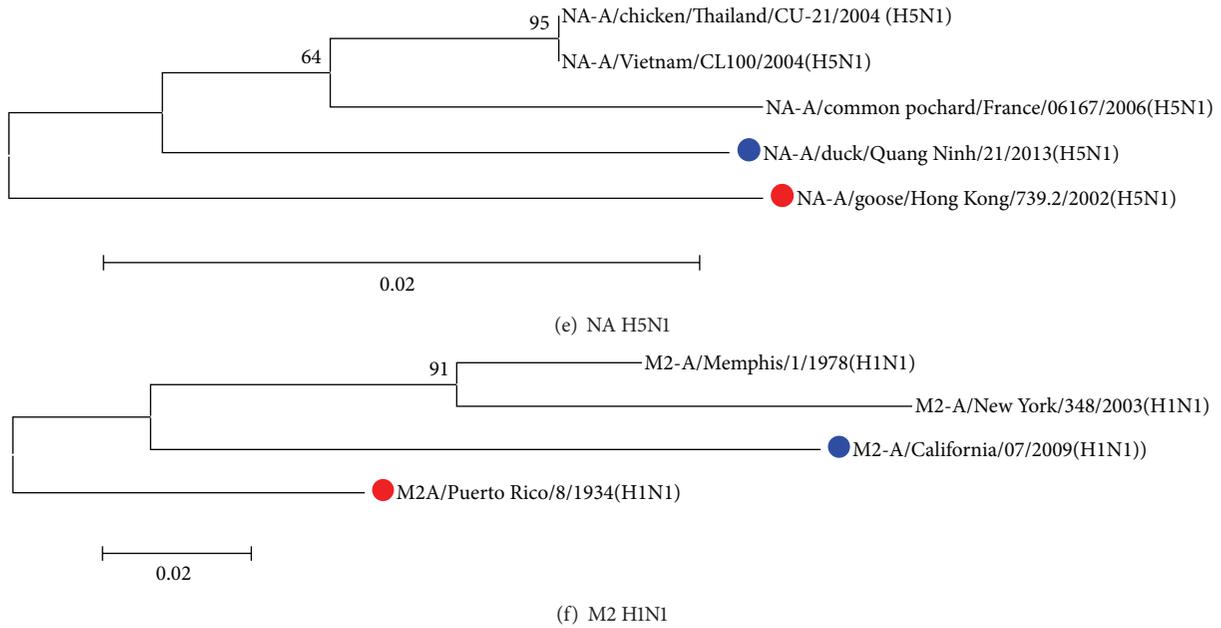
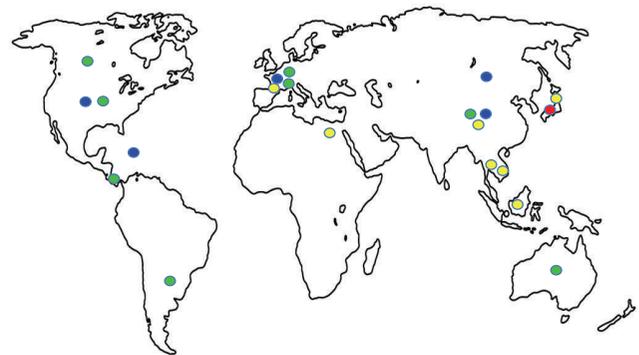


FIGURE 5: The red circle indicates the sequence in which the epitope was first identified. The blue circle shows the 2013-2014 season vaccine strain, in the case of H1N1 and H3N2, and the current strain, in the case of H5N1. The evolutionary history was inferred using the Neighbor-Joining method. The following branch shows the replication percentages with which the taxa associated in the bootstrap test. The trees are drawn to scale, with the longest of the branches in the same units of evolutionary distance used to infer the phylogenetic tree. The evolutionary distance was calculated using the Poisson method and is presented in substitution site units. The analysis was completed using the MEGA 6 program. Shown are (a) the optimum tree with a sum of branches of 0.43300672; (b) the optimum tree with a sum of branches of 0.30107033; (c) the optimum tree with a sum of branches of 0.20974655; (d) the optimum tree with a sum of branches of 0.29223355; (e) the optimum tree with a sum of branches of 0.07724998; (f) the optimum tree with a sum of branches of 0.28312403.

TABLE 4: PDB accession codes.

Subtype	Protein	PDB codes
H1N1	HA	2WR3
	NA	4B7J
	M2	2RLF
H2N2	HA	2WRD
	NA	4GZQ
H3N2	HA	2YP2
	NA	4GZQ
	NP	3TC6



- H1N1
- H2N2
- H3N2
- H5N1

FIGURE 6: Geographic distribution of consensus epitopes. The countries in which the isolates containing the consensus epitopes defined in this study have been obtained are shown on the map.

strain capable of crossing the species barrier and obtaining the capacity of being transmitted from human to human [57]. Thus, it is incredibly important to be prepared for the eventual introduction of this subtype with studies such as this one that determine the consistent recognition zones that can be used as targets in epitope vaccination techniques. With the current methodology for producing influenza vaccines using embryonated eggs, an emergency with a new strain would require 6 to 8 months to create a vaccine [58], as occurred in 2009.

Studying the highly recognized zones or those zones with the greatest numbers of epitopes within the IAV proteins over the years is related to site conservation; their remaining

constant makes it possible to identify the strains that will circulate in the coming years. Some of the consensus epitopes defined in this study are so constant that they have circulated for 91 years; these epitopes have been described in various studies conducted using sequences belonging to the 1918 strain and reported from 1983 (by Hackett and Cols) until

2012 (by Vergara and Cols). In the case of H2N2, there are epitopes that have been maintained for 44 years and have been described in studies completed from 1983 (by Lamb and Cols) until 2009 (by Babon and Cols). This study is the first to consider epitopes reported in strains that have circulated since 1918 to determine their conservation over such a long period of time.

It is not unusual that, for some IAV proteins, almost their entire sequence can be considered an epitope. Nevertheless, not all of these epitopes have the capacity to produce a protective response by inducing neutralizing antibodies, a fact that has been well verified in studies such as Gelder et al. in 1995 [37] and in 1998 [59], Alexander et al. in 2010 [43], and Rhee et al. in 2012 [40].

Of the 3,016 epitopes analyzed in this study, 18 consensus epitopes were identified, according to what was reported in the IEDB. Twelve of the epitopes were recognized exclusively by T lymphocytes [28, 31–35, 39, 43], while 6 represent overlapping antigens that are recognized by both B and T lymphocytes [29, 30, 32, 36–38, 40–42] (Table 3).

These epitopes were identified thanks to the fact that they have sequences of different sizes for the same site; thus, it was possible to determine the consensus sequence, that is, to eliminate the extremes from the reported sequences to determine, in an *in silico* manner, the minimum sequence that could be recognized by the immune system (Figure 3).

Another important aspect to take into consideration is how accessible or exposed these regions are for recognition by the immune system. In the case of the HA membrane protein, the majority of the consensus epitopes identified were located in the globular region (HA1); thus, we can infer that this region in general and, more specifically, the region close to the sialic acid-binding site show a greater degree of conservation than the stem region. These findings are different than those reported by Bulimo et al. in Kenya in 2012 [46] in a study that was completed during the 2007-2008 season; however, they support those findings reported by Iba et al. [60] in a study completed with the H3N2 strain with globular region epitopes with greater antigenic overlap than those found in the stem region. These findings emphasize the importance of conserving the cellular recognition region of the virus.

In the case of the NA protein, the consensus epitopes were identified far from the zanamivir-binding site, which would be expected, as this site is the site of greater selective pressure, resulting in a tendency for more modification. These findings coincide with those reported for the H1N1 subtype by Moscona in 2009 [61] and by Boivin in 2013 [62] and also for the H3N2 subtype by Tamura et al. in 2013 [63].

For the M2 protein, consensus epitopes were detected in the membrane and transmembrane regions, even though this is the Amantadine-binding site. Although resistance to ion channel blockers is less common for the M2 protein [64] than the resistance caused by neuraminidase (NA protein) inhibitors, it is hoped that the transmembrane region would have greater variability.

Another point to consider when searching for constant sites is the mutation rate for IAV proteins, which, contrary to what was expected, are 3-fold lower for the HA proteins of the H1N1 subtype than for HA proteins of the H3 and H5

subtypes. Similarly, the mutation rates are also 3-fold lower than that reported by Klein and colleagues [65] in 2014. This difference could occur because the periods of time evaluated were much shorter in those 3 cases, 41, 16, and 4 years, than the 91 years between the strains evaluated in this study. Another interesting fact is that the Klein study simply analyzed strains that circulated between 2009 and 2013 (from the pandemic), compared to our study, which analyzed strains from 1918 to 2009 (before the pandemic). Therefore, the mutation rate for the H1N1 subtype has apparently increased since 2009. This observation is reinforced if we consider that, between 1918 and 2009, the H1N1 subtype mutated due to genetic drift (i.e., only considering seasonal strains); yet, starting from the 2009 pandemic strain, modifications in its genome are the sum of both antigenic drift and antigenic shift (seasonal and pandemic strains). The same is true for the NA and NP proteins of the H1N1 subtype when comparing the findings from this study and those from Klein.

Finally, one aspect to explore is the reach of the vaccination or diagnostic strategies that could develop from these epitopes, as, currently, seasonal influenza vaccine design is based on strain circulation [66], independently for the northern [67] and southern hemispheres [68]. This fact increases the relevance of the consensus epitopes defined in this study, as, with the exception of the H2N2 subtype, all of the other subtypes have circulated on at least 3 continents, increasing their potential for use as therapeutic or diagnostic tools.

## 5. Conclusion

In conclusion, this study was able to identify 18 epitopes present in the HA, NA, NP, and M2 proteins of IAV that are, in accordance with previous studies, able to induce an immune response via T and B cells. These epitopes have remained constant for up to 91 years and have circulated on various continents. Nevertheless, because the analysis was completed in an *in silico* manner, it is necessary to demonstrate the potential of these findings experimentally in future studies.

## Conflict of Interests

The authors declare that there is no conflict of interests surrounding the publication of this paper.

## Acknowledgments

José Esteban Muñoz-Medina was the Consejo Nacional de Ciencia y Tecnología (National Science and Technology Council) grant holder for Grant no. 216252. The authors Eva Ramón-Gallegos and Alfonso Méndez-Tenorio are research assistants from the Sistema Nacional de Investigadores (National System of Researchers), from the Comisión de Operación y Fomento de Actividades Económicas del Instituto Politécnico Nacional (Operations and Economic Activities Development Commission of the National Polytechnic Institute), and from the Estímulos al Desempeño de los Investigadores (Researcher Achievement Incentive) program.

## References

- [1] T. Horimoto and Y. Kawaoka, "Influenza: lessons from past pandemics, warnings from current incidents," *Nature Reviews Microbiology*, vol. 3, no. 8, pp. 591–600, 2005.
- [2] P. Palese, "Influenza: old and new threats," *Nature Medicine*, vol. 10, supplement 12, pp. S82–S87, 2004.
- [3] W. W. Thompson, L. Comanor, and D. K. Shay, "Epidemiology of seasonal influenza: use of surveillance data and statistical models to estimate the burden of disease," *Journal of Infectious Diseases*, vol. 194, supplement 2, pp. S82–S91, 2006.
- [4] World Health Organization, *Fact Sheet 211*, World Health Organization, Geneva, Switzerland, 2009.
- [5] R. G. Webster, W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka, "Evolution and ecology of influenza A viruses," *Microbiological Reviews*, vol. 56, no. 1, pp. 152–179, 1992.
- [6] Q. Li, X. Sun, Z. Li et al., "Structural and functional characterization of neuraminidase-like molecule N10 derived from bat influenza A virus," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 46, pp. 18897–18902, 2012.
- [7] S. Tong, X. Zhu, Y. Li et al., "New world bats harbor diverse influenza A viruses," *PLoS Pathogens*, vol. 9, no. 10, Article ID e1003657, 2013.
- [8] Y. K. Choi, H. Ozaki, R. J. Webby et al., "Continuing evolution of H9N2 influenza viruses in Southeastern China," *Journal of Virology*, vol. 78, no. 16, pp. 8609–8614, 2004.
- [9] M. D. de Jong and T. T. Hien, "Avian influenza A (H5N1)," *Journal of Clinical Virology*, vol. 35, no. 1, pp. 2–13, 2006.
- [10] R. A. M. Fouchier, P. M. Schneeberger, F. W. Rozendaal et al., "Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 5, pp. 1356–1361, 2004.
- [11] Centers for Disease Control and Prevention, *Key Facts about Avian Influenza (Bird Flu) and Avian Influenza A (H5N1) Virus*, Centers for Disease Control and Prevention, Atlanta, Ga, USA, 2006.
- [12] Y. C. Hsieh, T. Z. Wu, D. P. Liu et al., "Influenza pandemics: past, present and future," *Journal of the Formosan Medical Association*, vol. 105, no. 1, pp. 1–6, 2006.
- [13] A. Rambaut and E. Holmes, "The early molecular epidemiology of the swine-origin A/H1N1 human influenza pandemic," *PLoS Currents*, 2009.
- [14] V. Shinde, C. B. Bridges, T. M. Uyeki et al., "Triple-reassortant swine influenza A (H1) in humans in the United States, 2005–2009," *The New England Journal of Medicine*, vol. 360, no. 25, pp. 2616–2625, 2009.
- [15] V. Trifonov, H. Khiabaniyan, B. Greenbaum, and R. Rabadan, "The origin of the recent swine influenza A(H1N1) virus infecting humans," *Euro Surveillance*, vol. 14, no. 17, 2009.
- [16] J. Chen and Y.-M. Deng, "Influenza virus antigenic variation, host antibody production and new approach to control epidemics," *Virology Journal*, vol. 6, article 30, 2009.
- [17] K. M. Neuzil, L. A. Jackson, J. Nelson et al., "Immunogenicity and reactogenicity of 1 versus 2 doses of trivalent inactivated influenza vaccine in vaccine-naïve 5-8-year-old children," *Journal of Infectious Diseases*, vol. 194, no. 8, pp. 1032–1039, 2006.
- [18] K. L. Nichol, P. M. Mendelman, K. P. Mallon et al., "Effectiveness of live, attenuated intranasal influenza virus vaccine in healthy, working adults: a randomized controlled trial," *Journal of the American Medical Association*, vol. 282, no. 2, pp. 137–144, 1999.
- [19] N. J. Cox and K. Subbarao, "Influenza," *The Lancet*, vol. 354, no. 9186, pp. 1277–1282, 1999.
- [20] S. A. Plotkin and W. A. Orenstein, *Vaccines*, Elsevier, 4th edition, 2004.
- [21] D. C. Ekiert and I. A. Wilson, "Broadly neutralizing antibodies against influenza virus and prospects for universal therapies," *Current Opinion in Virology*, vol. 2, no. 2, pp. 134–141, 2012.
- [22] J. Du, T. A. Cross, and H.-X. Zhou, "Recent progress in structure-based anti-influenza drug design," *Drug Discovery Today*, vol. 17, no. 19–20, pp. 1111–1120, 2012.
- [23] S. C. Gilbert, "Advances in the development of universal influenza vaccines," *Influenza and Other Respiratory Viruses*, vol. 7, no. 5, pp. 750–758, 2013.
- [24] T. Ben-Yedidia and R. Arnon, "Epitope-based vaccine against influenza," *Expert Review of Vaccines*, vol. 6, no. 6, pp. 939–948, 2007.
- [25] N. V. Kaverin, I. A. Rudneva, E. A. Govorkova et al., "Epitope mapping of the hemagglutinin molecule of a highly pathogenic H5N1 influenza virus by using monoclonal antibodies," *Journal of Virology*, vol. 81, no. 23, pp. 12911–12917, 2007.
- [26] T. T. Wang, G. S. Tan, R. Hai et al., "Vaccination with a synthetic peptide from the influenza virus hemagglutinin provides protection against distinct viral subtypes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 44, pp. 18979–18984, 2010.
- [27] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar, "MEGA6: molecular evolutionary genetics analysis version 6.0," *Molecular Biology and Evolution*, vol. 30, no. 12, pp. 2725–2729, 2013.
- [28] E. A. James, A. K. Moustakas, D. Berger, L. Huston, G. K. Papadopoulos, and W. W. Kwok, "Definition of the peptide binding motif within DRB1<sup>1401</sup> restricted epitopes by peptide competition and structural modeling," *Molecular Immunology*, vol. 45, no. 9, pp. 2651–2659, 2008.
- [29] X. Ge, V. Tan, P. L. Bollyky, N. E. Standifer, E. A. James, and W. W. Kwok, "Assessment of seasonal influenza A virus-specific CD4 T-cell responses to 2009 pandemic H1N1 swine-origin influenza A virus," *Journal of Virology*, vol. 84, no. 7, pp. 3312–3319, 2010.
- [30] Z. Majer, S. Holly, G. K. Toth et al., "Mapping the intersubunit region of influenza virus hemagglutinin: comparative CD and FTIR spectroscopic studies on multiple antigenic peptides," *Archives of Biochemistry and Biophysics*, vol. 322, no. 1, pp. 112–118, 1995.
- [31] K. A. Richards, F. A. Chaves, and A. J. Sant, "Infection of HLA-DR1 transgenic mice with a human isolate of influenza A virus (H1N1) primes a diverse CD4 T-cell repertoire that includes CD4 T cells with heterosubtypic cross-reactivity to avian (H5N1) influenza virus," *Journal of Virology*, vol. 83, no. 13, pp. 6566–6577, 2009.
- [32] J. A. B. Babon, J. Cruz, L. Orphin et al., "Genome-wide screening of human T-cell epitopes in influenza A virus reveals a broad spectrum of CD4<sup>+</sup> T-cell responses to internal proteins, hemagglutinins, and neuraminidases," *Human Immunology*, vol. 70, no. 9, pp. 711–721, 2009.

- [33] W. Zhong, P. A. Reche, C.-C. Lai, B. Reinhold, and E. L. Reinherz, "Genome-wide characterization of a viral cytotoxic T lymphocyte epitope repertoire," *The Journal of Biological Chemistry*, vol. 278, no. 46, pp. 45135–45144, 2003.
- [34] Y. S. Hahn, C. S. Hahn, V. L. Braciale, T. J. Braciale, and C. M. Rice, "CD8+ T cell recognition of an endogenously processed epitope is regulated primarily by residues within the epitope," *Journal of Experimental Medicine*, vol. 176, no. 5, pp. 1335–1342, 1992.
- [35] J. A. B. Babon, J. Cruz, F. A. Ennis, L. Yin, and M. Terajima, "A human CD4<sup>+</sup> T cell epitope in the influenza hemagglutinin is cross-reactive to influenza A virus subtypes and to influenza B virus," *Journal of Virology*, vol. 86, no. 17, pp. 9233–9243, 2012.
- [36] B. Morrissey, M. Streamer, and K. M. Downard, "Antigenic characterisation of H3N2 subtypes of the influenza virus by mass spectrometry," *Journal of Virological Methods*, vol. 145, no. 2, pp. 106–114, 2007.
- [37] C. M. Gelder, K. I. Welsh, A. Faith, J. R. Lamb, and B. A. Askonas, "Human CD4<sup>+</sup> T-cell repertoire of responses to influenza A virus hemagglutinin after recent natural infection," *Journal of Virology*, vol. 69, no. 12, pp. 7497–7506, 1995.
- [38] D. C. Jackson, X.-L. Tang, L. E. Brown, J. M. Murray, D. O. White, and G. W. Tregear, "Antigenic determinants of influenza virus hemagglutinin. XII. The epitopes of a synthetic peptide representing the C-terminus of HA1," *Virology*, vol. 155, no. 2, pp. 625–632, 1986.
- [39] L. Y.-H. Lee, D. L. A. Ha, C. Simmons et al., "Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals," *Journal of Clinical Investigation*, vol. 118, no. 10, pp. 3478–3490, 2008.
- [40] J. W. Rhee, D. Kim, B. K. Park et al., "Immunization with a hemagglutinin-derived synthetic peptide formulated with a CpG-DNA-liposome complex induced protection against lethal influenza virus infection in mice," *PLoS ONE*, vol. 7, no. 11, Article ID e48750, 2012.
- [41] S. Khurana, A. L. Suguitan Jr., Y. Rivera et al., "Antigenic fingerprinting of H5N1 avian influenza using convalescent sera and monoclonal antibodies reveals potential vaccine and diagnostic targets," *PLoS Medicine*, vol. 6, no. 4, Article ID e1000049, 2009.
- [42] M. Roti, J. Yang, D. Berger, L. Huston, E. A. James, and W. W. Kwok, "Healthy human subjects have CD4<sup>+</sup> T cells directed against H5N1 influenza virus," *Journal of Immunology*, vol. 180, no. 3, pp. 1758–1768, 2008.
- [43] J. Alexander, P. Bilsel, M.-F. del Guercio et al., "Universal influenza DNA vaccine encoding conserved CD4<sup>+</sup> T cell epitopes protects against lethal viral challenge in HLA-DR transgenic mice," *Vaccine*, vol. 28, no. 3, pp. 664–672, 2010.
- [44] W. I. Beveridge, R. Webster, and W. Bean, *Evolution and Ecology of Human Influenza Viruses. Influenza: The Last Great Plague*, Heinemann, London, UK, 1977.
- [45] J. C. Gaydos, F. H. Top Jr., R. A. Hodder, and P. K. Russell, "Swine influenza a outbreak, Fort Dix, New Jersey, 1976," *Emerging Infectious Diseases*, vol. 12, no. 1, pp. 23–28, 2006.
- [46] W. D. Bulimo, R. A. Achilla, J. Majanja et al., "Molecular characterization and phylogenetic analysis of the hemagglutinin 1 protein of human influenza a virus subtype H1N1 circulating in Kenya during 2007-2008," *Journal of Infectious Diseases*, vol. 206, no. 1, pp. S46–S52, 2012.
- [47] M. Y. Chowdhury, S.-K. Seo, H.-J. Moon et al., "Heterosubtypic protective immunity against widely divergent influenza subtypes induced by fusion protein 4sM2 in BALB/c mice," *Virology Journal*, vol. 11, no. 1, article 21, 2014.
- [48] M.-C. Kim, Y.-N. Lee, E.-J. Ko et al., "Supplementation of influenza split vaccines with conserved M2 ectodomains overcomes strain specificity and provides long-term cross protection," *Molecular Therapy*, vol. 22, no. 7, pp. 1364–1374, 2014.
- [49] Y.-T. Lee, K.-H. Kim, E.-J. Ko et al., "New vaccines against influenza virus," *Clinical and Experimental Vaccine Research*, vol. 3, no. 1, pp. 12–28, 2014.
- [50] M. Zheng, J. Luo, and Z. Chen, "Development of universal influenza vaccines based on influenza virus M and NP genes," *Infection*, vol. 42, no. 2, pp. 251–262, 2014.
- [51] Centers for Disease Control and Prevention, *Influenza aviar en humanos*, 2014.
- [52] D. J. Alexander, "A review of avian influenza in different bird species," *Veterinary Microbiology*, vol. 74, no. 1-2, pp. 3–13, 2000.
- [53] I. H. Brown, "Summary of avian influenza activity in Europe, Asia, and Africa, 2006–2009," *Avian Diseases*, vol. 54, supplement 1, pp. 187–193, 2010.
- [54] K. S. Li, Y. Guan, J. Wang et al., "Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia," *Nature*, vol. 430, no. 6996, pp. 209–213, 2004.
- [55] J. S. M. Peiris, W. C. Yu, C. W. Leung et al., "Re-emergence of fatal human influenza A subtype H5N1 disease," *The Lancet*, vol. 363, no. 9409, pp. 617–619, 2004.
- [56] World Health Organization, *Cumulative Number of Confirmed Human Cases for Avian Influenza A/(H5N1) Reported to WHO, 2003–2014*, World Health Organization, Geneva, Switzerland, 2014.
- [57] Y. Watanabe, Y. Arai, T. Daidoji et al., "Characterization of H5N1 influenza virus variants with hemagglutinin mutations isolated from patients," *mBio*, vol. 6, no. 2, Article ID e00081-15, 2015.
- [58] Center for Disease Control and Prevention, *How flu vaccine made*, 2015.
- [59] C. Gelder, M. Davenport, M. Barnardo et al., "Six unrelated HLA-DR-matched adults recognize identical CD4<sup>+</sup> T cell epitopes from influenza A haemagglutinin that are not simply peptides with high HLA-DR binding affinities," *International Immunology*, vol. 10, no. 2, pp. 211–222, 1998.
- [60] Y. Iba, Y. Fujii, N. Ohshima et al., "Conserved neutralizing epitope at globular head of hemagglutinin in H3N2 influenza viruses," *Journal of Virology*, vol. 88, no. 13, pp. 7130–7144, 2014.
- [61] A. Moscona, "Global transmission of oseltamivir-resistant influenza," *The New England Journal of Medicine*, vol. 360, no. 10, pp. 953–956, 2009.
- [62] G. Boivin, "Detection and management of antiviral resistance for influenza viruses," *Influenza and other Respiratory Viruses*, vol. 7, supplement 3, pp. 18–23, 2013.
- [63] D. Tamura, H. T. Nguyen, K. Sleeman et al., "Cell culture-selected substitutions in influenza A(H3N2) neuraminidase affect drug susceptibility assessment," *Antimicrobial Agents and Chemotherapy*, vol. 57, no. 12, pp. 6141–6146, 2013.
- [64] P. K. C. Cheng, A. P. C. To, T. W. C. Leung, P. C. K. Leung, C. W. C. Lee, and W. W. L. Lim, "Oseltamivir- and amantadine-resistant influenza virus A (H1N1)," *Emerging Infectious Diseases*, vol. 16, no. 1, pp. 155–156, 2010.

- [65] E. Y. Klein, A. W. R. Serohijos, J.-M. Choi, E. I. Shakhnovich, and A. Pekosz, "Influenza A H1N1 pandemic strain evolution—divergence and the potential for antigenic drift variants," *PLoS ONE*, vol. 9, no. 4, Article ID e93632, 2014.
- [66] Centers for Disease Control and Prevention, *Composición de la Vacuna Estacional de Influenza*, Centers for Disease Control and Prevention, Atlanta, Ga, USA, 2014.
- [67] Centers for Disease Control and Prevention, *Selección de Cepas para a Temporada 2013-2014 Hemisferio Norte*, Centers for Disease Control and Prevention, Atlanta, Ga, USA, 2014.
- [68] Center for Disease Control and Prevention, *Selección de cepas para la temporada 2013-2014 hemisferio sur*, 2014.



**Hindawi**  
Submit your manuscripts at  
<http://www.hindawi.com>

