

A Bioinformatics Study of Complete Amino Acid Sequences' Changes of Hemagglutinin Antigen of H1N1 Influenza Viruses in GenBank From Year 2006 to 2013 in Iran

Ali Farhangi,¹ Bahram Goliaei,^{1*} Kaveh Kavousi,¹ Abbas Ashtari,² Mohammad Ali Bayatzadeh,² and Ali

Pourbakhsh²

¹Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

²Razi Vaccine and Serum Research Institute, Karaj, Alborz, Iran

*Corresponding author: Bahram Goliaei, Professor at Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran. Tel: +98-2166498672, E-mail: goliaei@ibb.ut.ac.ir

Received 2016 December 14; Revised 2017 January 18; Accepted 2017 January 24.

Abstract

Background: Influenza is a major cause of morbidity and mortality worldwide. Each year, influenza viruses cause epidemics by evading pre-existing immunity through mutations in major surface glycoprotein hemagglutinin, which helps in attachment of the viral strain on the host cell surface. Due to high mutation rate, only currently circulating strains should be used in the vaccines.

Objectives: The present study aimed at analyzing a dataset of complete amino acid sequences of HA to assess the extent of diversity among circulating strains of Iran, during years 2006 to 2013, and studying important amino acid changes as well as changes in predicted ligand binding sites that could enhance viral performance.

Methods: 110 sequences from 17 provinces were downloaded, edited, and classified. The alignment of sequences and creation of phylogenetic trees and similarity matrices were done using bioinformatics software, such as MEGA6.0, BioEdit, DNAsisMAX, and DNAsstar. Web-based analyses including SWISS-MODEL, Phyre2, and 3DLigandSite were used for evaluation of the second and third protein structures and prediction of ligand binding sites.

Results: The results showed that 2009 was an important transition year, which classified the selected isolates into two different distinct groups. This shows the importance of changes made during possible mutations in the genomic structure of the virus, which have made it antigenically different from the previous years. This pandemic strain became dominant in the next years, and has been used as a standard vaccine strain from 2010 onwards.

Conclusions: The results of this study can shed further light on better understanding of the antigenic evolution of H1N1 influenza viruses and can be useful for epidemiological studies.

Keywords: Antigenic Variation, Influenza A virus H1N1 Subtype, Hemagglutinins, Sequence Alignment, Phylogeny, Point Mutation, Computational Biology

1. Background

Influenza is a highly contagious acute viral disease of the respiratory tract that has caused disease in humans since ancient times. It is a major cause of morbidity and mortality worldwide. Patients develop acute respiratory disease symptoms with headache, high fever, myalgia, nausea, and malaise.

The world health organization (WHO) has estimated that around 1 billion cases of seasonal influenza infection occur each year with around 3 to 5 million cases of severe illness and 300 000 to 500 000 deaths (1). It has been estimated that there were around 50 million human deaths due to infection with the 1918 pandemic virus, and about 575 400 deaths due to the 2009 pandemic (2, 3). The H1Na viruses belong to the *Orthomyxoviridae* family that in-

cludes RNA viruses with segmented negative-sense single stranded genome. This family includes three genera of influenza viruses named type A, B, and C (4). The H1N1 viruses are classified within type A. Influenza A viruses are further classified to subtypes based on the antigenic properties of the external glycoproteins Hemagglutinin (HA) and Neuraminidase (NA) (5). The virus is enveloped and from the surface of the envelope extends the two transmembrane glycoproteins HA and NA (6) which are the main targets of the host humoral immune responses. These two antigens are the most variable antigens that cause protective immunity. Hemagglutinin serves as the viral receptor-binding protein and mediates fusion of the virus envelope with the host cell membrane (7). Neuraminidase is responsible for assisting in virus cell entry and the release and spread of progeny virions (8-10).

1.1. Hemagglutinin and Antigenic Diversity

Antigenic variation is a significant feature in influenza viruses and they are notorious for antigenic variation and the accompanying frequent recurrence of epidemics. In other words, changes in proteins of the virus, specially HA and NA would cause variants that are new in the population. Two sorts of antigenic variations occur: antigenic drift and antigenic shift (11). Antigenic drift, which is a minor antigenic change, is a gradual change of the genome acquired by point mutations. Hemagglutinin is one of the most important antigens for inducing protective immunity in the host and shows the greatest variation. The HA molecule is the most frequently affected by antigenic drift, especially antigenic sites, which are exposed to immunological pressure.

The antigenic mutations in the gene codified by the HA protein cause changes in the structure of the glycoprotein, resulting in virus strains that can no longer be neutralized by previous host antibodies. This could result in viruses that are able to replicate more efficiently and mutants that could be transmitted more easily. This phenomenon is observed frequently in human influenza viruses, as a result of the selective pressure imposed by the use of vaccines. The antigenic shift, which is a major antigenic change, can occur through one of these three mechanisms: 1) genetic reassortment when two different influenza virus strains infect a single cell and results in a virus with new antigenic proteins; 2) direct transfer of whole virus from another species, and 3) re-emergence of a virus that may have caused an epidemic many years earlier. Antigenic shift can lead to completely new virus variants that might have a pandemic potential.

Continuous changes of HA and NA genes have made vaccine strain selection challenging and they must be selected annually, based on the currently circulating strains. Collecting, reporting, and analysis of epidemiologic data obtained from inspection of antigenic changes in each year is an important procedure, emphasized by the new WHO standards for global influenza surveillance, and could be useful in selection of the vaccine strains (12, 13).

2. Objectives

In general, studying changes and mutations in conserved positions and regions, especially those located in antigenic sites, having a potential for being the active site for binding with cell-receptors or form the structure of binding site in HA, are very important because these regions in a protein sequence usually have functional or structural values and can be used to develop targets for new drugs, infection control, treatment, and prediction for vaccine design.

This study aimed at analyzing the complete amino acid sequences of HA antigen of type A H1N1 influenza viruses available in GenBank from 2006 to 2013, in Iran, in order to define the prevalence library and study the trend of changes in HA sequences and to determine their relationship with changes of virus using bioinformatics methods. The relationship of gene alterations and the tertiary structure (3D model) with relevant regions in HA was also studied.

3. Methods

For collecting the required data and creating the data bank, the complete amino acid sequences (complete cds) of HA antigen of human H1N1 influenza virus, available in GenBank (<http://www.ncbi.nlm.nih.gov>) from 2006 to 2013, in Iran, were downloaded and saved in FASTA format. In order to obtain proper and identical sequences in terms of name and format for comparison and performing intended analyses, all downloaded sequences were edited using the DNASISMAX3.0 software (Hitachi, Pharmacia, Hitachi Software Engineering Company, Yokohama, and Japan) and classified on the basis of year of isolation. In total, 110 sequences from 17 provinces of Iran were downloaded, edited, and classified.

For evaluation and exact comparison of sequences, determination of conserved and variable regions, and similarities and differences in sequences, they were aligned and the phylogenetic trees and percentage similarity/divergence matrices were created. The BioEdit7.1 program was used to align all the sequences, based on the ClustalW method (14). The sequence alignments were analyzed using the Molecular Evolutionary Genetic Analysis (MEGA6.0) software (<http://www.megasoftware.net>) and the evolutionary distances between the strains were computed with the same software, using the maximum composite likelihood method (15, 16). Phylogenetic trees were constructed using the Megaline-Dnastar software (Expert Analysis Software for PC, Inc., USA) by the UPGMA method (unweighted pair group method with arithmetic means).

Using the percentage similarity/divergence matrix created by the BioEdit program in each year, the sequences with the highest and lowest similarity with other sequences of that year were selected as distinct sequences for comparison and analysis. Then web-based analysis including SWISS-MODEL (<http://swissmodel.expasy.org>) (17), Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2>) (18), and 3DLigandSite (<http://www.sbg.bio.ic.ac.uk/3dligandsite>) (19), were used for evaluation of secondary and tertiary protein structure and prediction/ determination of ligand

binding sites, type, and number of ligand(s) of distinct sequences of each year.

4. Results

The results of sequence alignments and phylogenetic trees (Figure 1), predicted secondary and tertiary structures (3D structures) of ligand binding sites of distinct sequences for each year and percentage similarity/ divergence matrices (data not shown and is available upon request) showed that in 2006 to 2008, the sequences were identical and there was no change in ligand binding sites, type, and number.

In 2009, with respect to the occurrence of H1N12009 pandemic, by studying the available similarities and differences in alignments of sequences and by investigating the phylogenetic tree of 2009 isolates (Figure 2), it was found that all influenza isolates are located in two entirely distinct and separated branches and form two different groups in terms of amino acid properties.

The remarkable point is that, the first group, in terms of amino acid sequences and gene positions, is similar to the sequences of 2006 to 2008 isolates and in some cases is identical, while the second group is similar to the sequences of 2010 to 2013 isolates and in some cases is identical. This similarity, in terms of alignment analyses of amino acid sequences and the structure of phylogenetic trees between the two different groups in 2009, shows the importance of changes in the genomic structure of the virus in this year, which were made during possible mutations in genome, likely by occurrence of frame shift or gene shuffling, causing different sequences and a new group. This new group became dominant in the following years, with significant amino acid sensible changes in HA genome that resulted in isolates with different bioinformatics and clinical properties. Moreover, it was found that, in 2009, the distinct sequence, which had the highest similarity (approximately 99%) with the other sequences of this year (ADJ18175) also had the highest similarity with the sequences of the following year's isolates (2010 to 2013), and the distinct sequence, which had the lowest similarity (approximately 77%) with the other sequences of 2009 (ADG23241), had the highest similarity with the sequences of previous years' isolates (2006 to 2008) (Figure 1). This shows the domination of 2009 pandemic strain from 2010 onwards.

By predicting and analyzing the secondary and tertiary structures in the amino acid sequences of 2009 and comparison of these structures, and analysis and comparison of regions and amino acid sequences in binding ligands and active sites of protein in molecular interactions between ligand and receptor, the differences between these

two groups in amino acid sequences analysis (bifurcation of phylogenetic tree) of 2009 isolates became more clear. It was found that in ligand-binding site 1 with 12 ligands, amino acid positions 239-240-241, in two distinct sequences of 2009 isolates were changed from Asp-Arg-Gln to Asp-Gln-Glu (Figure 3).

Since 2010 to 2013, sequences have been identical and there is no change in ligand binding sites, type, and number of ligands. This result shows that the predominant 2009 pandemic strain is also dominant in these years. From 2009 onwards, some changes in ligand binding sites 2 and 3 of distinct sequences of each year were observed compared to previous years from Thr255 in site 2 to Ile183-Lys256, and from Asp292-Asn303 in site 3 to Asn293-Asn304 (Table 1). There were also some changes in ligand binding sites, type, and number of ligands in 2012 in compare to previous and next years as follows: from Asp239-Gln240-Glu241 in site 1 to Lys236-Asp239-Arg240, and from Ile183-Lys256 with 6 ligands in site 2 to Ile183-Gly254-Lys256 with 7 ligands in site 2, and from Asn293-Asn304 with 4 ligands in site 3 to Lys39-Asn40 with 10 ligands in site 3 (Table 2).

Table 1. Changes in Ligand Binding Sites of Distinct Sequences from 2009 Onwards in Comparison to Previous Years

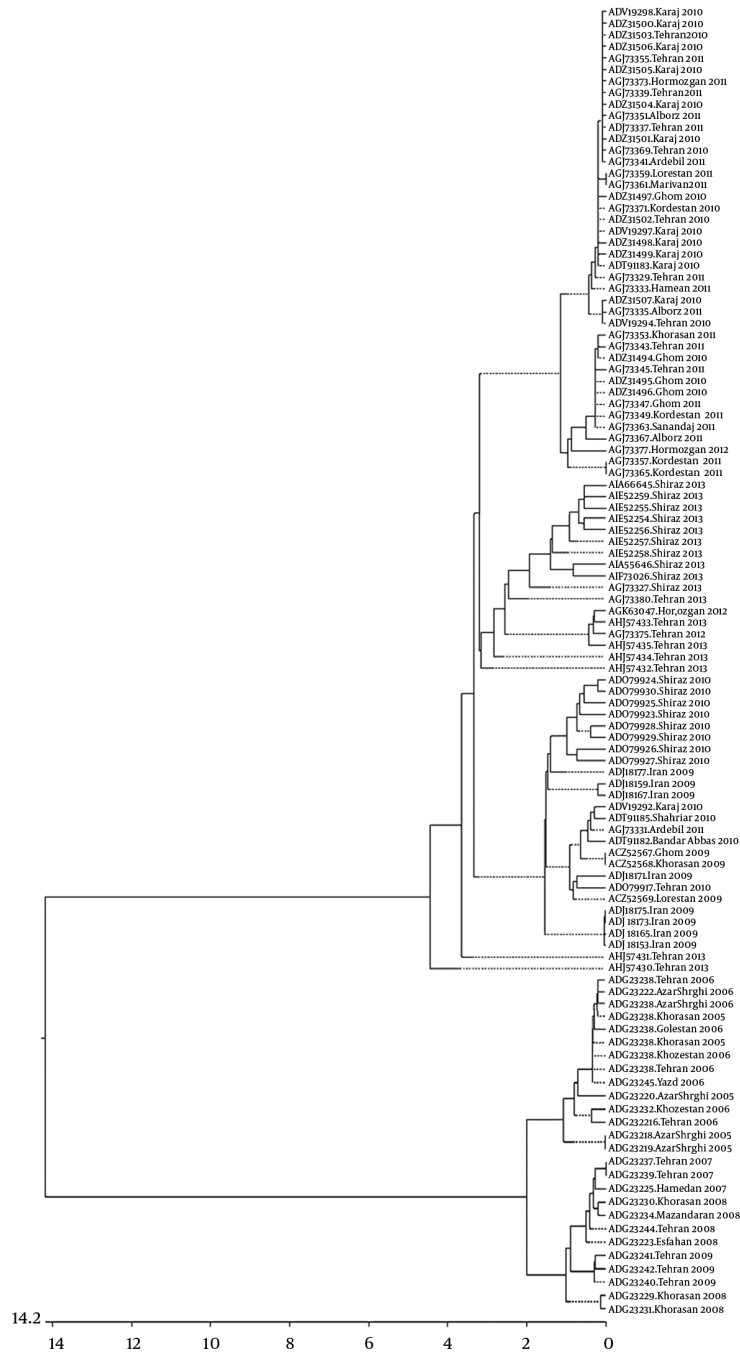
Ligand Binding Site	Year of Isolation	
	2006 - 2008	2009 - 2013
2	THR255	ILE183-LYS256
3	ASP292-ASN303	ASN293-ASN304

Table 2. Changes in Ligand Binding Sites, Types and Number of Ligands of 2012 Isolates of Iran in Comparison With the Next and Previous Years

Ligand Binding Site	2012	
	Change From	To
1	ASP239-GLN240-GLU241	LYS236-ASP239-ARG240
2	ILE183-LYS256 with 6 ligands	ILE183-GLY254-LYS256 with 7 ligands
3	ASN293-ASN304 with 4 ligands	LYS39-ASN40 with 10 ligands

In the next phase of this study, the sequences of different isolates were studied in terms of position and number of glycosylation sites (ASN- X- SER/ THR, in which X can be any amino acid except proline). The presence of glycosylation sites, which are commonly conserved, mask the protein surface from recognition by an antibody. Addition of glycans to the HA is an important mechanism contributing to antigenic drift and therefore sustained circulation of influenza A virus in the human population (20, 21). The

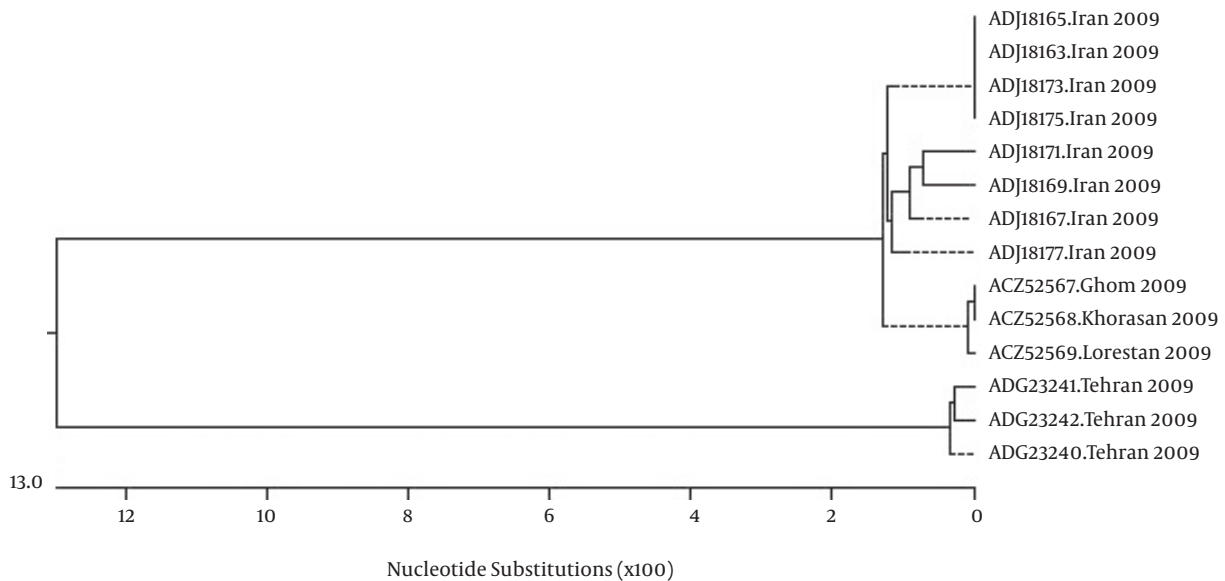
Figure 1. Phylogenetic Tree of the Hemagglutinin (HA) Gene by the UPGMA Method



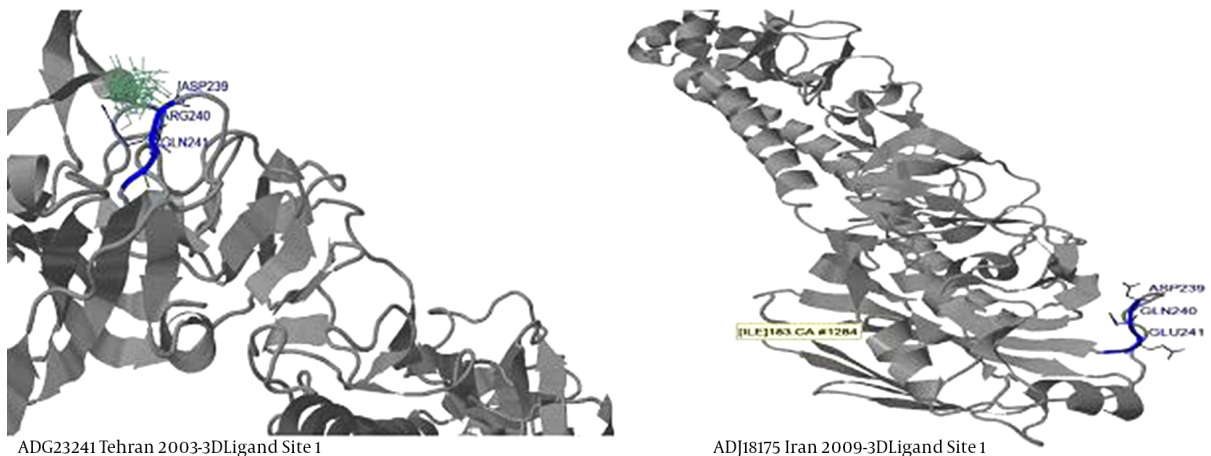
Evolutionary relationships of individual coding regions from influenza A viruses sampled in Iran during 2006 to 2013. Length of each branch pair represents the evolutionary distance between sequence pairs. Scale indicates the number of substitution. MegalignTM constructed phylogram (DNASTar, Madison, WI).

sequences of 2006 to 2008 isolates have 10 glycosylation sites at positions: 27, 28, 40, 71, 104, 142, 176, 303, 497, and 556, except some cases (Table 3). The sequences of 2009

to 2013 isolates have 8 glycosylation sites at positions: 27, 28, 40, 104, 293, 304, 498, and 557, except a few cases (Table 3), in which some changes (creation/ deletion of glycosyla-

Figure 2. Phylogenetic Tree of the Hemagglutinin (HA) Gene by the UPGMA Method

Evolutionary relationships of individual coding regions from influenza A viruses sampled in Iran during 2009. Length of each branch pair represents the evolutionary distance between sequence pairs. Scale indicates the number of substitution. Megalign™ constructed phylogram (DNASar, Madison, WI).

Figure 3. Predicted 3D Structure of the Ligand Binding Site of 2009 Distinct Sequences

3D Ligand site visualization of the protein structure with the residues, which form part of the binding site, is colored blue. The ligands that formed the cluster used for the prediction are also displayed as wireframes.

tion sites) have occurred. This implies a change in the number and location of glycosylation sites of 2009 pandemic strain, which became dominant in the following years.

Position 240, which is located in receptor binding site, is one of the important positions in HA (22). This position is conserved in sequences of 2006 to 2008 isolates, yet, from 2009 onwards, Q240R mutation has occurred

in some isolates. It has been found that this mutation greatly increases infectivity of virus without affecting its antigenicity.

Positions that have the potential for being the active site to bind with cell-receptors, and form the structure of binding site in HA have been listed in Table 4 (23, 24). By evaluation and analysis of sequence alignments of each

Table 3. Changes (Creation/Deletion) of Glycosylation Sites in Different Positions (Creation of Glycosylation Site at Positions 71, 142, and 177 in 2009 and Deletion at Positions 142 in 2006 and 2009, 293 in 2009, 104 in 2010 and 2013, and 304 in 2013)^a

Year	Positions										
	27	28	40	71	104	142	176/7	293	303/4	497/8	556/7
2006	C	C	C	C	C	NHT→KHT	C	-	C	C	C
2007	C	C	C	C	C	C	C	-	C	C	C
2008	C	C	C	C	C	NHT→KHT	C	-	C	C	C
2009	C	C	C	KCN→NCS	-	NHD→NHT	KLS→NLS	NTT→DAK	C	C	C
2010	C	C	C	-	NGT→KWT	-	-	C	C	C	C
2011	C	C	C	-	C	-	-	C	C	C	C
2012	C	C	C	-	C	-	-	C	C	C	C
2013	C	C	C	-	NGT→KGT	-	-	C	NTS→NPS	C	C

^a C, Conserved; -, No glycosylation site position.

year, the conservation and variability of these positions were determined (Table 4). It was found that, positions S153/ 4, N155/ 6, T168/ 9, H209/ 10, E211/ 12, and S274/ 5 are conserved in the period of 2006 to 2008, yet, from 2009 onwards, they were changed to S154P, N156A, T169V, H210Q, E212A, and S275E, respectively. In other words, in 2009, these positions were changed as mentioned above and these changes have remained conserved from 2010 to 2013. It was found that in 2009, in some cases, this mutation had not occurred.

Another investigated conserved region is the fusion peptide region. This region is critical for viral fusion function and is composed of 23 amino acid residues, including several large hydrophobic and several glycine residues interspersed throughout the sequence. The fusion peptide sequence GLFGAIAGFIEGGWTGMVDGWYG is extremely well conserved in H1N1 viruses (24). It was found that, the peptide fusion is well conserved in sequences of 2006 to 2013 isolates of Iran.

Another important position that was found and detected in pandemic 2009 H1N1 viruses, yet, not in previous H1N1 human viruses was V169. This position is associated with receptor binding specificity of HA (24). It was found that this position is conserved from 2009 to 2013 and may acquire V169T change in the human population in the near future. In 2009 this mutation was observed in Iran. It has been found that, D239G substitution in the HA of the pandemic 2009 H1N1 virus would increase the severity of disease. This substitution is associated with change of receptor binding affinity (25). It was found that, this position is almost conserved in 2009 to 2013 isolates except a few cases in each year, in which this substitution has occurred.

Molecular and phylogenetic analyses were also performed to reveal the relationship of changes of each year's isolates to the vaccine strains (eg, A/California/07/2009 (H1N1), which is the standard vaccine strain from 2010 onwards). It was found that all changes and obtained results

of each year are in compliance with and observed in standard vaccine strains of that year and the seasonal isolates turn out to be closely related to the corresponding vaccine strains. This means that studying and identifying antigenic changes of the virus in each year can be used for the prediction of changes in the future.

5. Discussion

Continuous changes of HA genes (antigenic variability) have always made influenza vaccine development challenging. Change in amino acid(s) of antigenic sites of influenza virus HA as well as other sites such as glycosylation sites and receptor/ligand binding sites may affect the potential of virus infection and spread within and between hosts. The study of changes in above-mentioned sites in available sequences of different viruses in a specified time period using conventional methods in bioinformatics can provide useful and important information about the trend of virus changes for virologists and health authorities to monitor the emergence of new influenza variants and to avoid the emergence of a virus with potential to cause a pandemic. It is of considerable importance to detect new antigenic changes occurring in HA protein when updating vaccine compositions.

In this study, by computational investigation of sequences isolated between 2006 and 2013 from Iran and inspection of the amount of phylogenetic similarity among them and by determination of binding sites and relevant ligands of HA antigen, it was found that the amount of similarity in sequence alignment analyses from 2006 to 2008 on one side, and from 2010 to 2013 on the other side is significant and in 2009, a set of two series of sequences of each of the two groups is detectable.

It was found that 2009 is a transition year, which almost classified the isolates to two groups. The 2006 to 2008 isolates and 2010 to 2013 isolates and each group had

Table 4. Conservation and Variability in the Positions as Potential Active Sites That Form the Structure of the Binding Site in Hemagglutinin ^a

Year	Positions										
	Y115	G147/8	S153/4	N155/6	W166/7	T168/9	L207/8	Y208/9	H209/10	E211/12	S274/5
2006	C	C	C	C	C	C	C	C	C	C	C
2007	C	C	C	C	C	C	C	C	C	C	C
2008	C	C	C	C	C	C	C	C	C	C	C
2009	C	C	I54P	I56A	C	I69V	C	C	I210Q	I212A	I275E
2010	C	C	I54P	I56A	C	I69V	C	C	I210Q	I212A	I275E
2011	C	C	I54P	I56A	C	I69V	C	C	I210Q	I212A	I275E
2012	C	C	I54P	I56A	C	I69V	C	C	I210Q	I212A	I275E
2013	C	C	I54P	I56A	C	I69V	C	C	I210Q	I212A	I275E

^a Conserved

similar changes/mutations. The 2009 phylogenetic tree was also divided to two distinct branches (due to the occurrence of the pandemic) with two groups of isolates. One group is similar with 2006 to 2008 isolates in terms of amino acid properties and changes/ mutations and the other group is similar with 2010 to 2013 isolates. This shows the importance of changes in the genomic structure of the virus in 2009 causing different sequences and a new group. This new different group and relevant mutations became dominant and stable in the next years, and the 2009H1N1 pandemic strain is used as a standard vaccine strain from 2010 onwards.

The results of our study are consistent with the results of others in terms of changes in number and location of glycosylation sites (20, 21), mutations/conservation in important positions such as Q240R (22), V169T, and D239G (24, 25), changes/conservation in active/binding sites (23, 24), and conservation in important domains such as fusion peptide (24, 25). Evaluation of changes in predicted ligand-binding sites, type and number of ligands was done in our study, which indicated a significant change in ligand binding site of 2009 isolates. This also shows the importance of 2009 as a transition year.

5.1. Conclusion

In conclusion, our findings suggest that the 2009 H1N1 pandemic strain might continue to mutate in its HA gene and create a new dominant strain with the potential to cause pandemics in the future. Therefore, conduction of this study is important to assess the diversity of circulating strains and to analyze the amino acid changes and third structure, which can change the function of the virus and its virulence or pathogenicity. It also emphasizes the importance of continuous monitoring of influenza A H1N1 virus strains (specially the A (H1N1) pdm09 strain) in different geographical areas to obtain better selection of the vaccine strain for upcoming seasons. The results obtained

in this study, can aid better understanding of the antigenic evolution of H1N1 influenza viruses and can be useful in epidemiological studies.

Acknowledgments

The authors would like to thank Dr. Azim Akbarzadeh and Dr. Dariush Norouzian, former and current head of NanoBiotechnology Department of Pasteur Institute of Iran for their kind support and advice.

Footnotes

Authors' Contribution: None.

Financial Disclosure: The authors had no financial interests related to the material in the manuscript and any other conflicts of interest.

Funding / Support: This research did not receive any specific grant from funding agencies in the public, commercial or non-profit sectors.

References

1. Organization WHO . Influenza. 2008.
2. Taubenberger JK, Morens DM. 1918 Influenza: the mother of all pandemics. *Emerg Infect Dis.* 2006;**12**(1):15–22. doi: [10.3201/eid1201.050979](https://doi.org/10.3201/eid1201.050979). [PubMed: [16494711](https://pubmed.ncbi.nlm.nih.gov/16494711/)].
3. Dawood FS, Iuliano AD, Reed C, Meltzer MI, Shay DK, Cheng PY, et al. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *Lancet Infect Dis.* 2012;**12**(9):687–95. doi: [10.1016/S1473-3099\(12\)70121-4](https://doi.org/10.1016/S1473-3099(12)70121-4).
4. Bouvier NM, Palese P. The biology of influenza viruses. *Vaccine.* 2008;**26** Suppl 4:D49–53. [PubMed: [19230160](https://pubmed.ncbi.nlm.nih.gov/19230160/)].
5. A revision of the system of nomenclature for influenza viruses: a WHO memorandum. *Bull World Health Organ.* 1980;**58**(4):585–91. [PubMed: [6969132](https://pubmed.ncbi.nlm.nih.gov/6969132/)].
6. Wang C, Takeuchi K, Pinto LH, Lamb RA. Ion channel activity of influenza A virus M2 protein: characterization of the amantadine block. *J Virol.* 1993;**67**(9):5585–94. [PubMed: [7688826](https://pubmed.ncbi.nlm.nih.gov/7688826/)].

7. Skehel JJ, Wiley DC. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem.* 2000;**69**:531-69. doi: [10.1146/annurev.biochem.69.1.531](https://doi.org/10.1146/annurev.biochem.69.1.531). [PubMed: [10966468](https://pubmed.ncbi.nlm.nih.gov/10966468/)].
8. Gottschalk A. Neuraminidase: the specific enzyme of influenza virus and *Vibrio cholerae*. *Biochim Biophys Acta.* 1957;**23**(3):645-6. [PubMed: [13426178](https://pubmed.ncbi.nlm.nih.gov/13426178/)].
9. Matrosovich MN, Matrosovich TY, Gray T, Roberts NA, Klenk HD. Neuraminidase is important for the initiation of influenza virus infection in human airway epithelium. *J Virol.* 2004;**78**(22):12665-7. doi: [10.1128/JVI.78.22.12665-12667.2004](https://doi.org/10.1128/JVI.78.22.12665-12667.2004). [PubMed: [15507653](https://pubmed.ncbi.nlm.nih.gov/15507653/)].
10. Palese P, Compans RW. Inhibition of influenza virus replication in tissue culture by 2-deoxy-2,3-dehydro-N-trifluoroacetylneuraminic acid (FANA): mechanism of action. *J Gen Virol.* 1976;**33**(1):159-63. doi: [10.1099/0022-1317-33-1-159](https://doi.org/10.1099/0022-1317-33-1-159). [PubMed: [978183](https://pubmed.ncbi.nlm.nih.gov/978183/)].
11. Carrat F, Flahault A. Influenza vaccine: the challenge of antigenic drift. *Vaccine.* 2007;**25**(39-40):6852-62. doi: [10.1016/j.vaccine.2007.07.027](https://doi.org/10.1016/j.vaccine.2007.07.027). [PubMed: [17719149](https://pubmed.ncbi.nlm.nih.gov/17719149/)].
12. Organization WHO . WHO global technical consultation: global standards and tools for influenza surveillance. 2011.
13. W. H. O. Writing Group , Ampofo WK, Baylor N, Cobey S, Cox NJ, Daves S, et al. Improving influenza vaccine virus selection: report of a WHO informal consultation held at WHO headquarters, Geneva, Switzerland, 14-16 June 2010. *Influenza Other Respir Viruses.* 2012;**6**(2):142-52. doi: [10.1111/j.1750-2659.2011.00277.x](https://doi.org/10.1111/j.1750-2659.2011.00277.x). [PubMed: [21819547](https://pubmed.ncbi.nlm.nih.gov/21819547/)] e1-5.
14. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series.* 1999;**41**:95-8.
15. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol.* 2011;**28**(10):2731-9. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121).
16. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol.* 2013;**30**(12):2725-9. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197).
17. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* 2003;**31**(13):3381-5. [PubMed: [12824332](https://pubmed.ncbi.nlm.nih.gov/12824332/)].
18. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;**10**(6):845-58. doi: [10.1038/nprot.2015.053](https://doi.org/10.1038/nprot.2015.053).
19. Wass MN, Kelley LA, Sternberg MJE. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acid Res.* 2010;**38**(Web Server):W469-73. doi: [10.1093/nar/gkq406](https://doi.org/10.1093/nar/gkq406).
20. Tate MD, Job ER, Deng YM, Gunalan V, Maurer-Stroh S, Reading PC. Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. *Viruses.* 2014;**6**(3):1294-316. doi: [10.3390/v6031294](https://doi.org/10.3390/v6031294). [PubMed: [24638204](https://pubmed.ncbi.nlm.nih.gov/24638204/)].
21. Schulze IT. Effects of glycosylation on the properties and functions of influenza virus hemagglutinin. *J Infect Dis.* 1997;**176** Suppl 1:S24-8. [PubMed: [9240690](https://pubmed.ncbi.nlm.nih.gov/9240690/)].
22. Wang W, Castelan-Vega JA, Jimenez-Alberto A, Vassell R, Ye Z, Weiss CD. A mutation in the receptor binding site enhances infectivity of 2009 H1N1 influenza hemagglutinin pseudotypes without changing antigenicity. *Virology.* 2010;**407**(2):374-80. doi: [10.1016/j.virol.2010.08.027](https://doi.org/10.1016/j.virol.2010.08.027). [PubMed: [20869738](https://pubmed.ncbi.nlm.nih.gov/20869738/)].
23. Hu W. Identification of highly conserved domains in hemagglutinin associated with the receptor binding specificity of influenza viruses: 2009 H1N1, avian H5N1, and swine H1N2. *J Biomed Sci Eng.* 2010;**03**(02):114-23. doi: [10.4236/jbise.2010.32017](https://doi.org/10.4236/jbise.2010.32017).
24. Sriwilaijaroen N, Suzuki Y. Molecular basis of the structure and function of H1 hemagglutinin of influenza virus. *Proc Jpn Acad Ser B Phys Biol Sci.* 2012;**88**(6):226-49. [PubMed: [22728439](https://pubmed.ncbi.nlm.nih.gov/22728439/)].
25. Tse H, Kao RY, Wu WL, Lim WW, Chen H, Yeung MY, et al. Structural basis and sequence co-evolution analysis of the hemagglutinin protein of pandemic influenza A/H1N1 (2009) virus. *Exp Biol Med (Maywood).* 2011;**236**(8):915-25. doi: [10.1258/ebm.2011.010264](https://doi.org/10.1258/ebm.2011.010264). [PubMed: [21727184](https://pubmed.ncbi.nlm.nih.gov/21727184/)].