Identification of sequence mutations affecting hemagglutinin specificity to sialic acid receptor in influenza A virus subtypes

Usman Sumo Friend Tambunan*, Ramdhan

Department of Chemistry, Faculty of Mathematics and Natural Science, University of Indonesia, Depok Campus, Depok 16424, Indonesia; Usman Sumo Friend Tambunan - Email: usman@ui.ac.id; *Corresponding author.

Abstract:
The attachment of the hemagglutinin protein of the H1N1 subtype of the pandemic influenza A virus to the sialic acid receptor Sia(α2-6)Gal has contributed to the ability of the virus to replicate in the human body and transmit among humans. In view of the pandemic caused by the replication and transmission of the H1N1 virus, more studies on the specificity of hemagglutinin towards sialic acid and how it affects the replication and transmission ability of this virus among humans are needed. In this study, we have applied sequence, structural and functional analyses to the hemagglutinin protein of the pandemic H1N1 virus, with the aim of identifying amino acid mutation patterns that affect its specificity to sialic acid. We have also employed a molecular docking method to evaluate the complex formed between hemagglutinin protein and the sialic acid receptor. Based on our results, we suggest two possible mutation patterns: 1) positions 190 and 225 from glutamic acid and glycine to aspartic acid (E190D in A/Brevig Mission/1/18 (H1N1), A/New York/1/18(H1N1) and A/South Carolina/1/1918(H1N1) and G225D in A/South Carolina/1/1918(H1N1), A/South Carolina/1/1918(H1N1), and A/Puerto Rico/8/34(H1N1)) and 2) positions 226 and 228 from glutamine and glycine to leucine and serine, respectively (Q226L and G228S in A/Guiyang/1/1957(H2N2), A/Kayano/57(H2N2), A/Aichi/2/1968(H3N2), A/Hong Kong/1/1968(H3N2) and A/Memphis/1/68(H3N2)) that can potentially contribute to the specificity of hemagglutinin to Sia(α2-6)Gal, thereby enabling the replication and transmission of virus within and among humans.

Background:
Influenza A virus is enveloped and contains eight segments of single stranded, negative sense RNA, which encodes 11 proteins (HA, NA, NP, M1, M2, NS1, NEP, PA, PB1, PB1-F2, PB2). The PB2, PB1, and PA encoded proteins form a polymerase complex for transcription, and are associated at one end of each gene segment [1]. The classification of influenza A virus is based on the antigenic properties of the hemagglutinin and neuraminidase glycoproteins, which are expressed on the surface of virus particle. Hemagglutinin is responsible for the binding of virions to host cell receptor and membrane fusion [2]. The host cell consists of a terminal sialic acid residue that is usually found in either an α2-3 or α2-6 linkage to galactose. Human influenza virus prefers binding to Sia(α2-6)Gal linked saccharide, whereas avian influenza virus prefers binding to Sia(α2-3)Gal [1].

There have been three influenza A pandemics in the 20th century. While the first pandemic caused by the H1N1 subtype of the influenza A virus resulted about 50 million deaths in 1918-1919, the subsequent pandemics caused by the H2N2 and H3N2 subtypes in 1957 and 1967 respectively led to one million deaths each [3]. The H5N1 subtype became a serious threat in 2003-2008, due to its high death rate, as well as its potential to cause widespread pandemic [4]. More recently, outbreaks of H1N1, also known as the Swine Flu, spreading rapidly across major parts of the world have resulted in the classification of the phenomenon by the World Health Organization's classification as the 21st century pandemic [5]. For a viral strain to become pandemic, three requirements must be fulfilled. Firstly, the viral strain should be able to enter the human body. Secondly, it should be capable of replicating within human hosts and finally, it should cause diseases that are easily-transmissible among humans [6]. Recently, there have been several exciting investigations of the relationship between the binding mode of the hemagglutinin protein and the pandemic ability of the influenza A viruses. One interesting finding is that Sia(α2-3)Gal forms multiple strong hydrophobic and hydrogen bond interactions with the hemagglutinin protein of H5N1 virus, whereas Sia(α2-6)Gal exhibits weak interactions to the same protein [7]. On the other hand, Miotto *et al.* identified an influenza A PB2 component that can adapt to enable human-to-human transmissions [8]. More importantly, it has been proposed that the ability of the hemagglutinin protein of the H1N1 influenza A virus strain to attach to Sia(α2-6)Gal instead of Sia(α2-3)Gal has enabled the replication and transmission of the virus among humans [9].

Based on the observation that all influenza A viruses prior to the pandemic H1N1 strain are highly pathogenic and not transmissible, we propose that specific amino acid mutations present in the hemagglutinin protein of the pandemic H1N1 strain serve to enable its replication and transmission ability within and among humans. In this study, we investigated this hypothesis using sequence, structural and functional analysis methods, with the aim of identifying possible amino acid mutations that affect the specificity of the H1N1 hemagglutinin protein to the sialic acid receptor.
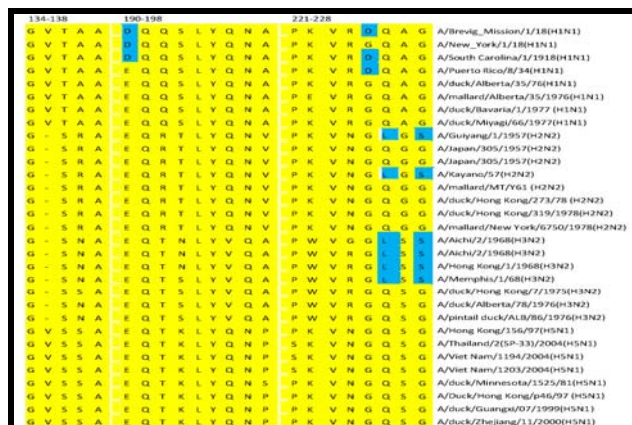
Figure 1: Sequence alignment of hemagglutinin of H1N1,H2N2, H3N2 and H5N1 from human and duck isolate.
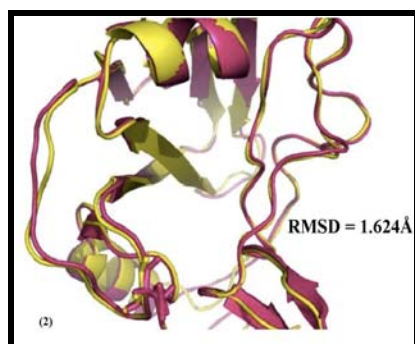

Figure 2: Superimpose between duck isolate (yellow) and human isolate (magenta) from H1N1.
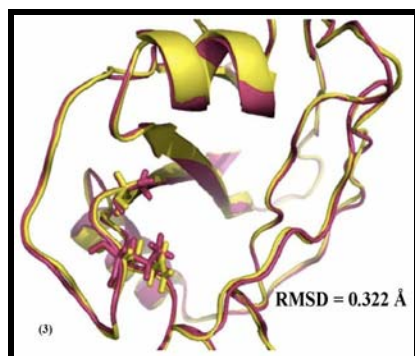

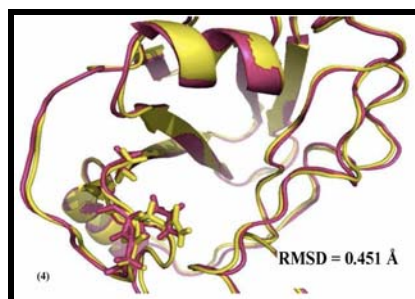**Figure 3:** Superimpose between duck isolate (yellow) and human isolate (magenta) from H3N2.

Figure 4: Superimpose between duck isolate (yellow) and human isolate (magenta) from H5N1.
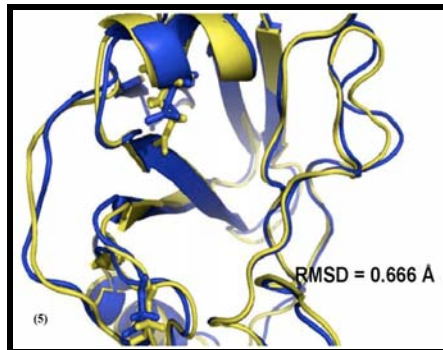
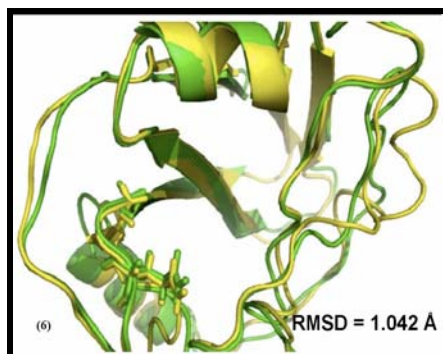**Figure 5:** Superimpose between H5N1 (yellow) and H1N1 (blue).

Figure 6: Superimpose between H5N1 (yellow) and H3N2 (green),
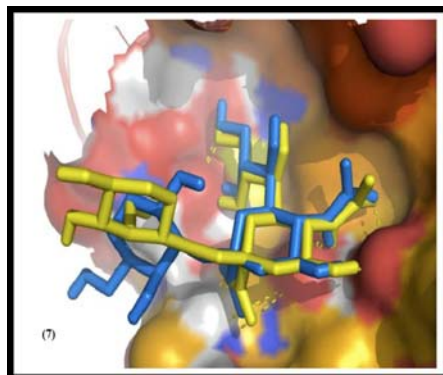

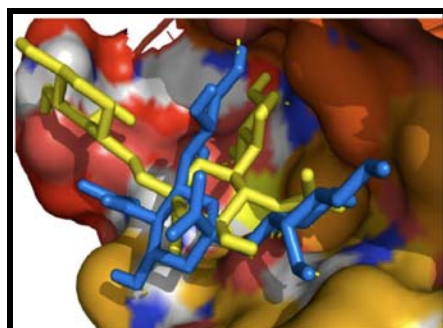**Figure 7:** Hemagglutinin 3HTT complex with Sia(α2-3)Gal [blue] and Sia(α2-6)Gal [yellow]

Figure 8: Hemagglutinin 1RD8 complex with Sia(α2-3)Gal [blue] and Sia(α2-6)Gal [yellow].

Methodology:
Data collection:
Amino acid sequences of hemagglutinin influenza A virus subtypes H1N1,H2N2, H3N2 and H5N1 isolated from humans and ducks were downloaded from the Influenza Virus Resources database created by the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genomes/FLU/Database.html). Only sequences isolated in pandemic years were collected, since hemagglutinin of these sequences binds specifically to Sia(α2-6)Gal. Sequences from ducks isolated in pandemic years were collected to serve as control protein sequences with binding specificity to Sia(α2-3)Gal **[10].**

Sequence mutation analysis:
ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html) **[11]** was used to perform multiple sequence alignment of the collected hemagglutinin protein sequence to identify specific mutations in the hemagglutinin receptor binding domain in human isolates compared to the control duck isolates. The sequence mutation analysis was limited to amino acids in the receptor binding domain which is located at the 130 loop (positions 134-138), 190 helix (positions 190-198) and 220 loop (positions 221-228). The positions are specific to H3 A/Aichi/2/68.

Structural Analysis:
The three dimensional structures of the hemagglutinin protein from different influenza A subtypes (both human and duck isolates), Sia(α2-6)Gal and Sia(α2-3)Gal were retrieved from the Protein Data Bank (http://www.rcsb.org/pdb/static.do?p=search/index.html). Hemagglutinin structures with 100 % similarity to the receptor binding domain of the hemagglutinin sequence were selected for each subtype **(Table 1, see supplementary material).**

The hemagglutinin structures of the duck and human isolates specific to each subtype were superimposed using Pymol v0.99 **[12]**. The Root Mean Square Deviation (RMSD) score, which represents similarity in the three-dimensional structure, were then calculated. RMSD value which are lower than 2 are deemed significant since it denotes two closely related protein structures **(Table 2, see supplementary material)**

Molecular docking:
The 3D structures of hemagglutinin proteins of the pandemic H1N1 strain isolated from human (PDB accession no: 1RD8) and from duck (PDB accession no: 3HTT) were docked to their respective sialic acids using Autodock 4.0. For each round of molecular docking, the hemagglutinin structure was defined as the receptor molecule, while structure of Sia(α2-3)Gal or Sia(α2-6)Gal were defined as the ligand. Both receptor protein and ligand molecules were prepared by removing all water molecules, adding hydrogen atoms and adding charges. The energies of the resultant docked complexes were minimized to relieve bad contacts using the conjugate gradient method in Vega ZZ 2.3.1 **[13].** The final structures were visualized and evaluated with Pymol v0.99**.**

Results:
Sequence mutation analysis:
The multiple sequence alignment (**Figure 1**) of the hemagglutinin receptor binding domains of influenza A H1N1, H2N2, H3N2 and H5N1 subtypes from duck and human isolates shows specific mutations in both positions 190 (in A/Brevig Mission/1/18(H1N1), A/New York/1/18(H1N1) and A/South Carolina/1/1918(H1N1)) and 225 (in A/South Carolina/1/1918(H1N1), A/South Carolina/1/1918(H1N1) and A/Puerto Rico/8/34(H1N1)). The sequences of all duck isolates contain glutamic acid in position 190 and glycine in position 225, whereas in the pandemic H1N1 human isolates, both positions have been substituted with aspartic acid. While both glutamic acid and aspartic acid at position 190 in the duck and human isolates possess highly similar physiochemical properties, the mutation at position 225 is one that involves a substation of hydrophobic glycine in the duck isolates with hydrophilic aspartic acid in the human isolates.

The H2N2 and H3N2 subtypes, on the other hand, exhibited a different mutation pattern between their corresponding duck and human isolates when compared to the H1N1 subtype. Mutations in H1N1 occurred in position 225, while mutations between the duck and human isolates in both the H2N2 and H3N2 subtypes occurred at positions 226 and 228 (in A/Guiyang/1/1957(H2N2), A/Kayano/57(H2N2) , A/Aichi/2/1968(H3N2), A/Hong Kong/1/1968(H3N2) and A/Memphis/1/68(H3N2)). Here, the amino acids glutamine at position 226 and glycine at position 228 in the duck isolates have been observed to mutate to leucine and serine in the human isolates, respectively. While the mutation at position 226 involves a substitution of hydrophilic glutamine with hydrophobic leucine, the mutation at position 228, on the other hand, has been observed to involve a reverse substitution of hydrophobic glycine with hydrophilic serine. Notably, no mutation has been observed in the influenza A H5N1 subtype, between the duck and human isolates, suggesting a possibility that the H5N1 subtype exhibits absolute specificity to Sia(α2-3)Gal.

Based on the multiple sequence alignment, we suggest two possible amino acid mutation patterns in the hemagglutinin receptor binding domain of influenza A viruses which can potentially contribute to its specificity towards Sia (α2-6)Gal. These mutation patterns are those occurring in positions 190 and 225 (E190 D and G225D) in the pandemic H1N1 strains and those in positions 226 and 228 (Q226L and G228S) in the H2N2 and H3N2 subtypes respectively. We propose that mutations that occur beside receptor binding domain can potentially affect substrate specificity, as well as antigenicity and pathogenicity.

Structural analysis of hemagglutinin receptor binding domain:
To visualize the structural differences between the duck and human isolates of each subtype, we superimposed the structures of one reference duck isolate and one reference human isolate for each subtype and measured the root mean square deviation (RMSD) value between the structures of these duck and human isolates.

The H1N1 (3HTT & 1RD8), H3N2 (1MQN & 1HGF), and H5N1 (1JSN & 2FK0) subtypes have been observed to possess RMSD values of 1.624Å (**Figure 2**), 0.322Å (**Figure 3**), and 0.451Å (**Figure 4**) between the duck and human isolates specific to each subtype respectively. The low RMSD values (<2Å) suggest that the receptor binding domain structures between the duck and human isolate in each specific subtype are closely related and have little structural differences. The results verify our hypothesis that the specificity of hemagglutinin to sialic acid can be influenced solely by amino acid changes in the receptor binding domain, highlighted in the multiple sequence alignment **[14]**.

The sequence mutation analysis has shown two possible amino acid mutation patterns that contribute to the recognition of Sia(α2-6)Gal in human isolates. We explored the proposed mutations at the three-dimensional level by visualizing the change in conformation in the receptor binding domain brought about by the mutations. We superimposed the structures of H5N1 and H1N1, as well as the H5N1 and H3N2 structures from human isolates. H5N1 was used as the reference point of comparison since it is the most threatening influenza virus. The RMSD values between H5N1 and H1N1 were determined to be 0.666Å (**Figure 5**), while the RMSD value between H5N1 and H3N2 were determined to be 1.042Å (**Figure 6**). The values imply that the structure differences between H5N1 and H1N1 are not obvious, while the structures of H5N1 and H3N2 are very closely related.

Molecular docking of hemagglutinin with sialic acid receptor:
Next, we performed molecular docking using Autodock 4.0 to dock the pandemic H1N1 hemagglutin structure to Sia(α2-3)Gal and Sia(α2-6)Gal respectively. The purpose of this analysis is to evaluate the quality of the two complexes and predict the specificity of hemagglutin to sialic acid. The quality of each complex was examined using the output score (Total Intermolecular Energy in kcal/mol) of AutoDock, which provides an estimation of ΔG° binding and inhibition constant. In general, a stable

complex is regarded to have been formed if the sialic acid residue can enter the receptor binding domain of hemagglutinin and interact with the important residues Asp190, Leu226, and Ser228.

Based on the general rule of thumb, the hemagglutinin structure 3HTT, which is an isolate from duck, possesses the ability to recognize Sia(α2-3)Gal and Sia(α2-6)Gal, since entry of both sialic acid residues to the receptor binding domain have been observed (**Figure 7**). The recognitions are in accordance with the minimum allowed Total Intermolecular Energy threshold, which further establishes the binding feasibility. However, the hemagglutinin isolate from human (PDB accession no: 1RD8) only exhibits specificity to Sia(α2-6)Gal, since only Sia(α2-6)Gal can enter the receptor binding domain (**Figure 8**). Results from molecular docking are consistent with the results of Glaser *et. al* that the mutation of single amino acid back to the avian consensus resulted in a preference for the receptor binding site of avian HA **[15].**

Discussion:
In this study, we have analyzed the sequences of the hemagglutinin protein across influenza A subtypes and identified mutation patterns occurring in positions 190 and 225 (E190 D and G225D) and positions 226 and 228 (Q226L and G228S) respectively, which can cause a change in chemical properties in the protein, ultimately affecting its binding specificity to sialic acid. The amino acid differences are investigated at a three-dimensional level via molecular docking methods to further validate the proposed change in binding specificity.

The results highlight the feasibility of an approach incorporating sequence, functional and structural analysis in predicting the effect of sequence mutations on the structure and function, specifically the human to human transmission capability, of the hemagglutinin protein in influenza virus. Integration approach as such employed in this study can be applied to facilitate the design of novel drugs. The high mutation rate of the virus necessitates the use of creative and intelligent methods to design novel treatment. In this regard, bioinformatics approaches represent one of the most feasible ways to cope with such a short and demanding time frame posed by these mutations. As such, influenza drug design in the future will be more dependent on *in silico* methods, due to the greater rate of discovery as compared to laboratory methods, which can better comply with the high mutation rate of the influenza virus. Nevertheless, bioinformatics approaches have to be ultimately validated and verified through laboratory experiments.

Conclusion:
Hemagglutinin is the primary determinant of host biochemical function for its role in host cell recognition and attachment. It has been established that hemagglutinin in human influenza viruses prefers Sia(α2-6)Gal linked saccharide, whereas hemagglutinin in avian influenza viruses prefers Sia(α2-3)Gal **[9]**. In our study, we identified two possible mutation

patterns in the receptor binding domain of the hemagglutinin protein of the pandemic H1N1 strain that contributes to the shift in recognition specificity from Sia(α2-3)Gal to Sia(α2-6)Gal, thereby potentially enabling the replication and transmissibility of influenza strains in turn. These mutation patterns identified include 1) mutations in positions 190 and 225 to aspartic acid (E190D in A/Brevig Mission/1/18(H1N1), A/New York/1/18(H1N1) and A/South Carolina/1/1918(H1N1) and G225D on A/South Carolina/1/1918(H1N1), A/South Carolina/1/1918(H1N1), and A/Puerto Rico/8/34(H1N1)) and 2) mutations in position 226 and 228 to leucine and serine (Q226L and G228S in A/Guiyang/1/1957(H2N2), A/Kayano/57(H2N2), A/Aichi/2/1968(H3N2) and A/Hong Kong/1/1968(H3N2), A/Memphis/1/68(H3N2)). To validate our results, we performed molecular docking analysis to predict hemagglutinin specificity by analyzing the computationally docked complex of hemagglutinin and sialic acid. The results show that such an approach can provide useful information for predicting mutations in the hemagglutinin protein in influenza A virus strains that could lead to the capability of human to human transmission and therefore, to cause pandemics.

References:
[1] LG Ross & K Subbarao. *Clin Microbiol Rev* **19**: 614 (2006) [PMID: 17041137]
[2] RG Webster *et al. Clin Microbiol Rev* **56**: 152 (1992) [PMID: 1579108]
[3] ED Kilbourne. *Emerg Infect Dis* **1**: 9 (2006) [PMID: 16494710]
[4] Ungchusak *et al. N Engl J Med.* **352**: 333 (2005) [PMID: 15668219]
[5] G Neumann *et al. Nature* **7249**: 931 (2009) [PMID: 19525932]
[6] DL Noah & RM Krug. *Adv Virus Res* **65**: 121 (2005) [PMID: 16387195]
[7] M Li & B Wang. *Biochem Biophys Res Commun.* **347**: 662 (2006) [PMID:16844080]
[8] O Miotto *et al. BMC Bioinformatics* **9**: s18 (2008) [PMID:18315849]
[9] TM Tumpey *et al. Science* **5812**: 655 (2007) [PMID: 17272724]
[10] A Gambaryan *et al. Arch Virol* **147**: 1197 (2002) [PMID:12111429]
[11] MA Larkin *et al. Bioinformatics* **23**: 2947 (2007) [PMID: 17846036]
[12] http://www.pymol.org
[13] A Pedretti, *et al. J C.A.M.D* **18**: 167 (2004) [PMID: 15368917]
[14] JF Gibrat *et al. Curr Opin Struct Biol* **6**(3): 377 (1996) [PMID: 8804824]
[15] Glaser *et al. J Virol* **79**: 11533 (2005) [PMID: 16103207]

Supplementary material:

Table 1: PDB accession numbers of hemagglutinin structures for each influenza A subtype used in this study.

Subtype	PDB ID	Isolate
H1N1	3HTT	Duck
	1RD8	Human
H3N3	1MQN	Duck
	1HGF	Human
H5N1	1JSN	Human
	2FK0	Human

Table 2: The threshold values for similarity determination of 3D structures, based on the Root Mean Square Deviation (RMSD) score.

RMSD (Å)	Structural Relationship
>12	Completely unrelated
7	Dubious relationship
5	May be structurally related
4	Good structural relationship
2	Closely related
1.5	Very closely related
0.8	Differences are not obvious
0.4	Essentially indistinguishable